PRACTICAL STATISTICS

WITH

FUNDAMENTALS OF THEORY

By
M ZIA-UD-DIN, M A , Ph D (Wales),
Head of the Statistics Department,
Panjab University, Lahore

1946
2nd Edition (Revised and Enlarged)

PUBLISHED BY
THE PUNJAB EDUCATIONAL PRESS
LAHORE

Price Rs 5/8/-

Printed and Published by
[illegible] M Malik at The Punjab Educational Press
35, Nabha Road, Lahore

PREFACE TO THE FIRST EDITION

Nowadays Statistics has become an important subject and commands its application in almost every science

This book has been written with a view to supply a practical knowledge of the commonly used statistical methods to the beginner although it does not claim to be a detailed treatise on Statistics

The methods have been clearly explained and illustrated with examples

This book covers the syllabuses in Statistics of the various examinations of the Punjab University, such as B A (Hons in Economics) M A B Sc and M Sc (Agriculture) and Commerce Examinations

I hope this book will prove to be useful for

(1) Statistical workers in general

(2) Students of the Punjab and other Universities

(3) Persons preparing for competitive examinations

A good number of Exercises have been given at the end of each Chapter *with answers* for practice

Books given in the Bibliography at the end have been consulted in preparing this book The Exercises have been mostly taken from various examinations, such as competitive examinations, examinations of the Universities Commerce examinations (Hailey College), class tests etc

The authar will be obliged for suggessions for the improvement of the book

Department of Statistics,
University of the Panjab,
Lahore
10th November 1943

M ZIA-UD-DIN

PREFACE TO THE SECOND EDITION

The first edition of Practical Statistics was finished very soon and the demand for the book has been great The book has been well appreciated by Professors Government officials students and the public interested in Statistics

The second edition is revised and enlarged The theory (economic as well as mathematical) has been added and the book is brought up to date The fresh additions are mathematical theory of interpolation summary of Bowley Robertson report list of statistical publications in India detailed form of Questionnaire mathematical proofs of theorems on probability and moments and Panjab University Question Papers for 1946 (Questions from other Universities and Competitive examinations are given in the Exercises)

In the Panjab University Statistics forms a subject for (1) Postgraduate certificate in Statistics examination (2) M Com examination and a paper for M A (Mathematics) M A (Economics) M.Sc (Agriculture) B Com B.A (Hons Econ) and B A (Mathematics B Course) This book will be found useful for all the examinations in Statistics

I am thankful to Professors and students for their kind suggestions

Department of Statistics
Panjab University
Lahore
2nd September 1946

ZIA UD DIN

CONTENTS

*Answers are g ven along with Exercises throughout the bool

CHAPTER

INTRODUCTION २९ १०-१९७८

Definition, Characteristics and Importance of Statistics

The word Statistics is used in the plural as well as in the singular sense Statistics in the plural are numerical facts systematically collected with some definite object in view in any field of inquiry, whatsoever, of observation or experiment. For example (*e g*) Statistics of, population, births and deaths, height and weight, income and expenditure imports and exports, crimes, morals, rainfall and temperature Railway passengers

Mere figures 60, 62, 65, 68, 70---are not statistics tl they are figures, but 60 seers, 62, 65, 68, 70 ..seers, weight of a class of students, will form statistics

The fundamental **characteristics** are .—

(1) Statistics should be expressed quantitatively. Qualitative words like good, fair, poor, young, healthy, will not be called statistics

(2) Statistics are aggregates that is made up of a numbei of individuals or cases A single sale, accident or birtl will not constitute statistics.

(3) Statistics must be prepared in a systematic manner keeping the given purpose or object in view as clear anc definite as possible

(4) Statistics are related to other facts, and should b homogeneous and comparable.

Statistical Method is a technique used to obtain, analyse summarise, compare and present the numerical data (or statistics).

Statistical methods consist of general rules, principles, graphical representation and formulæ applicable to all types of data

Statistics in the singular is a science which investigates the statistical methods and deals with their applications Statistics deals with (1) populations or aggregates of individuals, (2) Variations in population, (3) Reduction of bulky and incomprehensive data

Like all other sciences, statistics can be classified as (1) pure statistics, (2) applied statistics

Pure statistics or theoretical statistics is mathematical and deals with general theories, formulæ, equations and their derivation

Applied statistics deals with the application of statistical methods to concrete subject matter, such as, measurement of economic, commercial, social, agricultural, industrial, scientific and mental phenom-na, measurement of living organisms, study of vital and population movements and actuarial *principles*

Statistics plays an important part in every walk of life and has proved to be extremely useful in almost every line of scientific and economic inquiry

Economists, businessmen, industrial concerns, bankers, educationists, scientists, astronomers, navigators, insurance companies, railway traffic managements, public bodies, government's departments of public health, meterology, agriculture, industries, commerce, food, labour, post war *reconstruction* and planning, are largely benefited by the use of statistics and need statisticians

Limitations

Statistical method which is the only means for handling large masses of numerical data, is limited in its application to data which are reducible to quantitative form Statistical laws are true on the average and in the long run and do not study the individual constitution of a group They show approximate tendencies and estimates and they can be used even when experimental methods fail. Statistical methods should be intelligently and carefully used as their misuse may lead to ridiculous and unsatisfactory results. Fallacious conclusions will follow if the data supplied for statistical investigation are incomplete, unreliable and based on prejudiced collection and in such cases science of statistics is not to be blamed at all

Collection of Statistical Data

The following methods may be conveniently used to have a collection of statistical facts for an inquiry. These collected statistics should be as far as possible, reliable, accurate, clear, without ambiguity, unbiassed, comprehensive and complete for statistical investigations.

The unit of investigation determined should be definite, specific, homogeneous and stable.

The methods of collection may be briefly mentioned as · Method of

(1) Direct personal investigation or interview method ;

(2) Indirect investigation ; through correspondence.

(3) Sampling or representative data Sampling may be—

(*i*) deliberate or purposive, (*ii*) random, (*iii*) stratified, Sampling is described in detail later on (Chaptsr XI)

(4) Questionnaires, (for specimen see Appendix)

(*5*) Investigation on the basis of government publications, reports of the different departments of governments and states, gazettes, budgets, reports of banks and commercial concerns, research publications, trade and census reports and such other published documents.

Classification and Tabulation

After collection of data, the data should be classified and placed in a tabulated form, as described below

Variates —Any character which can vary in quality or in magnitude is called a variate, thus age, height, occupation, income, colour of the hair and examination marks are variates Some variates are measurable or quantitative, others are categoric or qualitative Age and income are quantitative variaties, colour of the hair and occupation, are categoric or qualitative, as they cannot be measured numerically

Classification may be made on either of the four bases·

(I) *Qualitative* —When the basis of distinction rests upon the differences in quality or condition An analysis of sales by reference to the kind of goods sold involves qualitative distinction

(2) *Quantitative, i.e.,* differences being in quantity. An analysis of sales according to differences in weight, volume or value of the goods involved in each transaction would be quantitative

(3) *Temporal.*—Involving the time at which the objects in question were measured, or the events in question oc An ana ysis of annual sales by weeks and months will invol temporal classification.

(4) *Spdtial or geographical*—Referring to the distribution of items in space or according to location, *e g* annual sales by geographical areas and places.

Classification may be simple and manifold. It be based on attributes or characteristics in respect of which some are similar, and others dissimilar

Classification according to one attribute, *e g*, deaf not deaf, blind, not blind, in which each class is divided into two sub-classes and no more, is said to be simple. If more than one attribute is noted, classification may be carried further giving rise to several classes and sub classes Such a classification will be called manifold classification

Class intervals and frequencies—Consider the following marks awarded out of 50, obtained by 30 students

3, 5, 8, 15, 25, 30, 16, 7, 35, 40, 49, 40, 30, 15, 14, 21, 23
22, 25, 27, 29, 32, 15, 1, 8, 9, 11, 14, 42, 43.

The data obtained as a result of observation o experiment in the original form are called 'ungrouped data When the data are split into groups or classes, they ar called grouped data. The marks given above, form an grouped data. These marks can be formed into groups or classes by first arranging them in ascending or descending order, as

1, 3, 5, 7, 8, 8, 9, 11 14, 14, 15, 15, 15, 16, 21, 22, 23, 25, 25, 27, 29, 30, 30, 32, 35, 40, 40, 42, 43, 49.

Such an arrangemeut in ascending or descending order is called an array

The data can be classified into groups as

Marks	*Number of Students.*	*Marks*	*Number of Students.*
1—5	3	31—35	2
6—10	4	36—40	2
11—15	6	41—45	2
16—20	1	46—50	1
21—25	5		
26—30	4		30

In the first group, 1 and 5 are the class limits, 1 is aid to be the lower limit and 5, the upper limit In each roup, 5 marks have been counted, so 5 is said to be the class interval' or the magnitude of the class-interval of the roup. The number of students securing grouped marks, gainst each group, is called the frequency of the class iterval or of the group The frequency of a variate ; the number of times it occurs. The data can also be lassified with 10 as class interval, counting the upper limit in he next group, as follows

Class intervals	*Frequencies.*
0—10	7
10—20	7
20—30	7
30—40	4
40—50	5
	30

The total number of frequencies is 30, and the upper limits is counted in the succeeding group

Some people write the classification, to avoid ambiguity, as

0 and under 10	7
10 ,, ,, 20	7
20 ,, ,, 30	7
30 ,, ,, 40	4
40 ,, ,, 50	5

The classification is, sometimes, written in the reverse order, taking the last group as first and so on

The following points may be kept in mind while classifying—

1 Class limits must be fixed with reference to the accuracy of the observation.

2. Suitable class intervals should be kept according to the size of the data It should not be so large as to make the grouped data, look very small, neither so small as to make it look unwieldy. The difference between the greatest and the least value of the data may be divided by the number of conveniently-sized groups to obtain approximately the class interval As far as possible, attempt should be made to have the class limits as integers and the class interval, itself also a whole number, to facilitate the application of further statistical methods.

3. The class intervals should be uniform as far as possible

4. There should not be indeterminate classes, that is the classes, the intervals of which are not defined, unless unavoidable, *e g*,

Age
Under 5
5—10

Age
10—20
Above 20

Here the first and last classes are indeterminate

5. For a fairly large data, the possible groups can be etween 10 and 25

Statistical Series

In order to analyse numerical data, it is necessary to rrange them systematically An arrangement of the data n a systematic order is called a distribution or series If the data be grouped according to magnitude or size, the eries formed is a **frequency distribution**, consisting of class ntervals and frequencies

Data grouped according to the time of occurence, form Time or Historical Series

If Data are grouped according to the geographic location, he resulting series is a spatial distribution

Continuous and discrete series A variate is said to be ontinuous when it passes from one value to the next by ndefinitely small gradations, *e g*, height and weight, where ve can have differences of small fractions A variate is aid to be discrete (or Integral or discontinuous) when there re gaps between one value and the next, *e g*, the number f children in a family, for families differ in size by one r more integers and not by fractions Continuous variates will form a continuous series and discrete variates ʼ discrete series

Tabulation—The classified data should be placed in form of Tables with rows and columns Tabulation is si and manifold or complex, according as classification is simple and manifold form A frequency distribution ʼ a frequency table The following general rules are to k kept in mind for tabulation

1. First make out a rough draft, but the tables dra should be accurate, attractive, neat and tidy

2. Avoid complicated tables Information of a h degree of complexity should be broken up into sections

3. The title should constitute a clean, concise an complete description of the material assembled in the table

4. Headings of the columns and rows should be concis and without any ambiguity

5. Columns and rows may be numbered to facilitat reference to the table.

6. The table should constitute a unit, self-sufficient an self-explanatory All explanation necessary for the interpre tation of the table should be included as integral part c the table or in the form of foot-notes

7. The tables should be so constructed as to be easi read and understood, its figures easily compared, and follow without unnecessary waste either of time or thought

The convenience of the person who needs the tab may also be consulted, and the sources of the data shou be given.

8 Card system and mechanical system of tabulation may be used when such a machinery is available

Nowadays Machines exist for tabulating as well as calcuating purposes

Errors

The divergence between the actual number and the stimate which is made either by approximation or by any ther method, is called an Error (it is not a mistake) Errors re absolute, relative, Biassed and unbiassed Statis ical errors may be due to incomplete and prejudiced nformation, in adequate sampling and in exact manipulation. f x represents the estimate, y, the true value, then the bsolute error e is $y-x$,

nd the relative error is $\frac{e}{x}\left[\text{or } \frac{e}{y} \text{ if the true value is taken}\right]$

If a quantity is such that its errors, are all in the same rection, the error is said to be biassed The greater e number of items the greater the error, that is why assed error is also called cumulative error If a quantity such that its errors tend to neutralise one another, the ror is said to be unbiassed or compensating Two important atistical errors namely standard error and probable error are scribed later on

CHAPTER II

MEASURES OF CENTRAL TENDENCY OR AVERAGES

The fundamental measures of central tendency or averages are—(1) Arithmetic Mean or Arithmetic average or simply Mean, (2) Median, and Quartiles, (3) Mode, (4) Geometric Mean, (5) Harmonic Mean, (6) Weighted average.

In this chapter we shall deal with (1) and (2).

The Arithmetic Mean is calculated as follows :—

(i) For ungrouped data, add all the given items and divide the sum by the number of items, *e g.* The Mean of Rs. 10, 20 and 30 will be $\frac{10+20+30}{3}=20$ Rs.

This is the simple Arithmetic average.

(ii) *Direct Method* for grouped data, *i.e*, when class intervals and frequencies are given. The formula is : Mean $=\frac{\Sigma fx}{n}$, where Σfx is the sum of the products of the central or middle or mean values of the groups and their corresponding frequencies, n is the total number of frequencies $=\Sigma f$. The symbol Σ is used for summation. S is also used in place of Σ (sigma)

Example.—Weekly wages (Rs. 5 internal)	*Central Values x*	*No. of employees or frequencies.*	$f \times x$
Rs 1—5	1+5 3	3	9
6—10	6+10 8	4	32
11—15	13	6	78
16—20	18 1164	1	18
21—25	23	5	115
26—30	28	4	112
31—35	33	2	66
36—40	38	2	76
41—45	43	2	86
46—50	48	1	48
		30	640

Here $\Sigma fx = 640$, $n = 30$, Mean $= \frac{640}{30} = 21\frac{1}{3}$ Rs

(iii) *Short cut method* —Take any Mean to be called a Provisional Mean, or Assumed Mean or Arbitrary origin, and find the deviations (differences) of the central values from the Provisional Mean The formula for Arithmetic Mean is then

$$\text{Arithmetic Mean} = \text{Provisional Mean} + \frac{\Sigma f \times d}{n}$$

where d, denotes the deviations of the middle values from the Provisional Mean Let us work out the above example by taking 13 as the Provisional Mean

x	f	d	$f \times d$
3	3	−10	−30
8	4	−5	−20
13	6	0	0
18	1	5	5
23	5	10	50
28	4	15	60
33	2	20	40
38	2	25	50
43	2	30	60
48	1	35	35
	30		250

Arithmetic Mean

$= 13 + \frac{250}{30} = 21\frac{1}{3}$

This gives the average wage The same result as by direct method

Any Provisional Mean may be taken, but as a convention, the middle value corresponding to the maximum frequency in the given distribution is to be taken as a Provisional Mean. The short cut method proves more useful in case of a large data, or if there are decimals, than the direct method For the sake of convenience, to avoid heavy multiplication, the magnitude of the class interval may be taken common out of the deviations In the above example 5 can be taken out common in column d, and then multiplied at the end by $\Sigma f \times d$, so formed

Advantages of Arithmetic Mean—(1) The Arithmetic average is the most commonly used average (2) It is easily calculated and understood and is the most generally recognised type of average (3) It utilises all the data in the groups

Disadvantage—Its value may be greatly distorted by the extreme values and, therefore, sometimes it may not be typical

Median Quartiles, Deciles and Percentiles—Consid an ungrouped data arranged in ascending or descending order, *i e* an arrayed data The middle item of the array is called the Median It is the central item which has as many items preceding as succeeding it When the number of items is odd, the median can be easily located *e g* If there are eleven items, the median will be represente by the 6th item (five items preceding and five, followin it) If n is the number of items, the median will b $\left(\frac{n+1}{2}\right)$th item When the number of items is even, ther will be two central values $\left(\frac{n}{2}\right)$-th and $\left(\frac{n}{2}+1\right)$th, item either of them can be taken as Median and the Mea of these two central values may be taken as the Media Value

For grouped data, *i e*, for a frequency distribution, th Median is calculated with the help of the formula

Median $=l+\frac{i}{f}\left(\frac{n}{2}-c\right)$ Where n is the total number of frequencies, $\frac{n}{2}$ the median number which will lie in a group whose lower limit Is l, i is the class interval of the Median group *i.e.*, in which the nsedian lies, and f its corresponding frequency, c denotes the cumulative frequency of the group preceding the Median group.

Example.—Let us work the median for the previous example.

Groups.	*Frequencies*	*Cumulative Frequencies.*
1—5	3	3
6—10	4	7 *i e.* (3+4)
11—15	6	13 *i e* (7+6)
16—20	1	14 *i e* (13+1}
21—25	5	19
26—30	4	23
31—35	2	25
36—40	2	27
41—45	2	29
46—50	1	30
	30	

The cumulative frequencies are 3, 7, (3+4), 13 (3+4+6), 14, 19, 23, 25, 2/, 29 and 30. $\frac{n}{2}=\frac{30}{2}=15$ lies in the group 21—25 This is the Median group, whose lower limit is 21, the given frequency corresponding to this group is 5 and i is also 5. Therefore, Median $=21+\frac{5}{5}(15-14)=22$.

For the Median number $\frac{n}{2}$ is used for continuous well as for for discrete series. But for discrete series $\frac{n+}{2}$ can be used when n is odd

Advantages and Disadvantages —The median is typical, when the central values of the series are closely grouped, and the array consists of terms quite close to other. It can be located by inspection and is not distorted by eztremes or unusual terms For the data of the type.

Frequencies

Below 5	---	
5—10	--	Median is a better average
10—15	---	than Arithmetic Mean
Above 15	—	

Median is not so familiar an average as the Arithmetic Mean.

In locating the Median, the items have to be arrayed which is not done in the case of Arithmetic average.

Quartiles, Deciles and *Percentiles* —Just as the median divides the distribution into two parts, the Quartiles divide it into four parts, Deciles into ten parts and the Percentiles into one hundred parts To determine the values of these measures the same process is used as for median except that we us $\frac{n}{4}$ for first quartile Q_1 for second quartile, $\frac{2n}{4}$, and for third quartile $Q_3, \frac{3n}{4}$. Thus $Q_1 = l + \frac{i}{f}\left(\frac{n}{4} - c\right)$ $Q_3 = l + \frac{i}{f}\left(\frac{3n}{4} - c\right)$

For discrete series when n is odd $n+1$ can be used in place of n

For deciles we can use $\frac{n}{10}$ for first decile, $\frac{2n}{10}$ for second and so on

For Percentiles we may use $\frac{n}{100}$ for first Percentile $\frac{2n}{100}$ for second and so on and proceed exactly as for median

Besides these measures of comparison we have also Quintiles and Octiles which divide the distribution into five and eight parts respectively The rest of the formulæ and process is the same as for Median for all these measures

In the above example $Q_1 = 11 + \frac{5}{6}\left(\frac{30}{4} - 7\right) = 11\frac{5}{12}$, as $\frac{30}{4}$ lies in the group 11—15

Exercise I

1 How will you proceed to conduct an economic inquiry of your own native place?

2 Prepare Questionnaires for (1) your own college (2) big factory or firm (3) well known Bank

3 Draw Table for the data given at the end of the Exercise I

4 Draw Tables to show the distribution of population of your Province by (1) age, sex and literacy (2) Sex and occupations

5 Form frequency table of the following taking class intervals as 2, 2 5 and 3 respectively

Rupees 1, 7, 4, 2 8, 3 5, 4 2 3 5 7 2, 5, 3 4, 15, 25, 17 2 19 3, 19, 27, 26 4, 29 1, 30 2, 14 6, 2 3, 1 5, 29 13, 10 45, 13 7, 14 9, 2 7, 3 5, 19 3, 20 9, 17 5, 12 8, 1 9, 3 9

6. Find the Arithmetic Mean and Median of the following observations 22, 24, 20, 25, 21, 18, 23, 22, 20, 22, 20, 22, 23, 25, 21, 24, 21, 22, 24, 23, 22, 23, 21, 22, 21, 23.

Ans Mean 21·96 and Median 22.

7. Form a frequency distribution of the following data giving the index numbers of 60 commodities in a certain year and find the value of the Mean and the Median 76, 78, 81, 85, 86, 87, 89, 90, 91, 94, 95, 96, 96, 98, 97, 99, 99, 100, 100, 101, 101, 101, 102, 104, 104, 104, 105, 105, 106, 107, 108, 108, 108, 109, 109, 110, 110, 111, 112, 113, 113, 114, 115, 115, 116, 116, 117, 117, 118, 119, 120, 121, 122, 123, 124, 125, 128, 129, 134

(M A 1942 Punjab University)

(*Ans* 106 85) 107 22

8. Given,

Height in inches (x)	73,	72	71,	70,	69	68,	67,	66,	65
Men (f)	2	4	6	10	11	[illegible]	5	4	1

Calculate the mean height.

Ans 69 78

9. Given

Variate,	[illegible]	19,	18	17	16,	15,	14,	13,	12	11
Frequency	1	2	4	8	11	10	7	4	2	1

Find the Mean variate (1) by taking 11 as the origin (2) 15 as Zero (i e as Pro Mean) and verify by the direct method

Ans 15·54

10. Given the following frequency distribution, calculate the Arithmetic Mean

Monthly Wages Rs Rs	*Workers*	*Monthly Wages* Rs Rs	*Workers*
12 5—17 5	2	37 5—42 5	4
17 5—22 5	2[illegible]	42 5—47 5	6
22 5—27 5	19	47 5—52 5	1
27 5—32 5	14	52 5—57 5	1
32 5—37 5	3		

(*M Sc Agriculture 1943*)

Ans Rs 27 85

11 (*a*) Find the median quartiles, 8th decile and 56th percentile for the following distribution

Class Intervals Rs	*Frequencies*	*Class Intervals* Rs	*Frequencies*
1—2 99	6	11—12 99	16
3—4 99	53	13—14 99	4
5—6 99	85	15—16 99	4
7—8 99	56		
9—10 99	21	Tota	245

Hint —In such decimal classes consider class interval, a whole number, in this case the interval is 2 and groups as class interval 2

1—3
3—5 for computation

Ans $6\ 49$ $Q_1=5\ 05$, $Q_3=8\ 4$
$D_8=8\ 86$ *and* $P_{56}=6\ 8$

(b) Given

Rs		Rs	
4— 7 999	4	28—31 999	22
8—11 999	16	32—35 999	10
12—15 999	46	36—39 999	2
16—19 999	68	40—43 999	2
20—23 999	58	44—47 999	0
24—27 999	32	48—51 999	1

Calculate the Arithmetic Average

Ans 20 6

√12 Calculate the median, the lower Quartile and the upper Quartile for the following frequency distribution of the number of marks obtained by 49 students in a class —

Marks obtained	*No of Students*	*Marks obtained*	*No of Students*
5 − 10	5	25 − 30	5
10 − 15	6	30 − 35	4
15 − 20	15	35 − 40	2
20 − 25	10	40 − 45	2

(*Punjab University B A Hons 1942*

Ans 19 6 15 41, 25 75

√13 Find the median and the first Quartile

Amount of wages		*Number of workers so receiving such Rate of wages*
Not exceeding 10 shillings	..	50
Over 10s but not exceeding 12s		70
Over 12s „ „ „ 14s.	..	60
Over 14s „ „ „ 16s	..	81
Total	..	261

Hint—Take the median number as $\frac{261+1}{2}=131$

and for $Q_1=\frac{261+1.}{4}$

Ans 12s 4 4 pence $Q_1=$ *10s. 5 2d.*

14 Calculate the median

(*a*) x Rs. 10, 8, 6, 4, 2.
frequecy 1, 4, 6, 4, 1

(*b*) x Rs. 20, 40, 60, 80
f 10, 50, 30, 10.

(*c*) x 10, 12, 14, 16, 18, 24.
f 2, 5, 6, 4, 2, 1

(*d*) x 3, 5, 7, 9.
f 200, 400, 300, 100.

Hint.—First of all put x into class intervals, so as to have x as the middle values and then proceed in the ordinary way.

For (*d*) Class intervals are

γ	x	
2−4	3	3
4−6	5	8
6−8	7	15 ✓
8−10	9	24

Ans. (*a*) *6,* (*b*) *46,* (*c*) *14,* (*d*) $5\frac{7}{2}$

15. Find the Median and Quartiles for the follo frequency distribution.

	f	
Rs. 12, 8 ans.—Rs. 17, 8 ans.	4	4
„ [illegible] „ — „ [illegible] „	44	48
„ „ — „ [illegible] „	38	86
„ [illegible] „ — „ [illegible] „	28	114
„ „ — „ [illegible] „	6	120
„ [illegible] „ — „ [illegible] „	8	128
„ [illegible] „ — „ [illegible] „	12	140
„ „ — „ [illegible] „	2	142
„ 52, 8 „ — „ 57, 8 „	2	144
	144	

Ans. Median $= Rs.\ 25,\ 10\frac{10}{19}$ *ar*

$Q_1 \quad = Rs.\ 21,\ 2\frac{2}{11}$ *ans.*

$Q_3 \quad = Rs\ \ 31,\ 8\frac{6}{7}$ *ans*

16. The following table gives the number of males a emales in U. P. in 1921. Calcutta the average age of mal and females. Calculate

Age	*Males* (*in lakhs*).	*Females* (*in lakhs*)
0 − 10	61	58
10 − 20	49	38
20 − 30	40	38
30 − 50	60	54
50 − 80	23	28

Ans. Males $25\ \frac{110}{233}$

Females $26\frac{87}{108}$ ×

17 The frequency distribution below gives the cost of production of sugar-cane in different holdings, obtain the Arithmetic Mean.

	Frequency.		Frequency
2—6	1	18—	52
6—	9	22—	36
10—	21	26—	19
14—	47	30—34	3

(*Indian Audit and Account Service Exam 1941*)

Ans 19 2127

18 Calculate the values of the median and the two uartiles for the following.

Limits for percentage recovery of sugar on cane	*Factories in India (1935—36).*	
8'0—8 2	2	2
8'2—	5	7
8 4—	4	11
8 6—	11	22
8'8—	11	33
9' —	11	44 ✓
9'2—	13	57
9 4—	10	67
9'6—	7	74
9 8—	6	80
10' —	3	83
10 2—	1	84
10'4—10 6	1	85
	85	

(*M. A. 1943 Punjab University*)

Ans. Median, 9 18. Q_1 =8'78, Q_3 =9'55

✓19 The chest measurements of 10,000 men are given as follows —

Inches —33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48

Men —6, 35, 125, 338, 740, 1303, 1810, 1940, 1640, 1120, 600, 222, 84, 30, 5, 2.

Calculate the mean

(*M A 1941, Aligarh University*) *Ans* 39 835

✓20 The following table gives the distribution of the male and female population of a certain area in India Find the mean age, median age, upper and lower quartile ages.

Age groups.	*Males*	*Females*
0—9	2756	2787
10—19	2124	2032
20—29	1677	1724
30—39	1481	1485
40—49	1021	1022
50—59	610	579
60—69	245	269
70—79	67	78
80—89	16	20
90—99	3	4
	10,000	10 000

(*I C S. 1936*) *Ans Males* 25 649 20 71, 9 07, 36 37
Females 23 774, 21 05, 8 97, 36 44.

21 Calculate the mean and median for the following distribution

Weights of boys in a certain class, 100—104, 105—109,
Number 4 14
110—114, 115—119, 120—124, 125- 129, 130—134,
60 138 206 298 380
135—139, 140—144, 145—149, 150—154, 155—159,
450 500 430 260 128
160—164, 165—169, 170—174.
66 28 12 =2974

22 The following table gives the marks obtained by a batch of 15 candidates in a certain examination in History Politics and Economics In which subject is the level of knowledge of candidates highest ? Give reasons

Roll No	*History*	*Politics*	*Economics*
1	42	46	35
2	24	20	25
3	38	41	44
4	35	45	50
5	30	25	15
6	45	54	57
7	58	47	55
8	50	36	40
9	40	30	20
10	62	61	64
11	55	50	42
12	54	63	60
13	52	45	62
14	47	56	54
15	43	58	52

(*B A Hons 1945*) *Ans* (*Economics*)

23

Age groups	*Population in United Kingdom in millions*	*of India*
0—10	24	358
10—15	20	222
15—20	18	157
20—25	16	145
25—30	14	161
30—40	27	257
40—50	25	184
50—60	19	120

Compare the average age *Ans* U K 27 203
(*B Com Punjab 1945*). *India* 24 24.

24 The following table shows the frequency distributic of yield of wheat in maunds per acre in 998 irrigated field selected at random in the province of Punjab

Limits in Mds	No of fields
0—4	45
4—8	184
8—12	281
12—16	228
16—20	155
20—2	77
24—28	22
28—32	5
32—36	1

Calculate the average yield per acre

(C St Exam and M A 1945) *Ans 12 4C*

25 Make a frequency table having grades of with class intervals of two annas each from the following dat of daily wages received by 30 labourers in a certain factor and then compute the average daily wage paid to a labourer

Daily wages in annas

14, 16, 16, 14, 22, 13, 15, 24, 12, 23, 14, 20, 17, 21, 1 18, 19, 20, 17, 16, 15, 11, 12, 21, 20, 17, 18, 19, 22, 23.

(B A Hons 1945) *Ans. Rs 1, 2a*

26 The frequency distribution according to age of group of persons is as follow —

Age group	*No in the group*
0— 5	4
5—15	12
15—25	13
25—35	11
35—45	12
45—55	8
55—65	4
65—75	1

8/2 to

1164

(B A Hons 1945

Calculate the Median *Ans* 28 6 ✓

27 Calculate the Arithmetic mean for

Monthly Income Rs	Labourers.
12—16,	2
16—20,	8
20—24,	10
24—2	12
28—32,	15
32—36,	20
36—40,	12
40—44,	10
44—4	8
48—52,	6
52—56,	4
56—60	1

Ans. 33 81

(Hyderabad University B A. 19

28. The table shows the age distribution of married mal‘s according to samp e census of 1941 in the Baroda late

ge	0—5,	5—10,	10—15,	15—20,	20—25
umber of arried females	3	31	410	1809	2446
	25—30,	30—35,	35—40,	40—45	
	2223	1723	1292	963	
	45 —50,	50—55,	55—60,	60—65,	
	762	531	317	156	
	65—70,	70—75			
	59	37			

Calculate the median age of married females and also : two quartiles

Ans. 28'78 , 21 91 , 38 58.

(Indian Audit & Accountants Serv ce Exam, 1942).

29 Calculate the Quartiles for the following frequency tribution of weights of a certain class of people :—

Weights in poundf 100—105, 105—110

Number of persons 5, 10, 15, 65, 40, 32 ,

170 —175

44, 35, 40, 29, 30, 25, 15, 10, 8

Ans $120\frac{23}{32}$, $147\frac{93}{116}$

(Indian Audit & Acctt. Exam 1945)

30 Compile the statistical data contained in the llowing paragraph in tabular form —

The United States Bureau of Foreign and Domestic ommerce presented, in the December 1937 " Monthly

Summary of Foreign Commerce ' data of exports of United States merchandise and of imports for consumption (not including imports for purposes of re export), segregated into "economic classes and for various years Comparing 1936 and 1937, the total value of exports was $2 418,969,000 in 1936 and $3,294,916,000 in 1937, while the total value of imports for consumption was $2 423,977,000 in 1936 and $3,012,487 000 in 1937 Crude materials exported in 1936 amounted to $668,168,000, or 27 6 per cent of the total value of exports for that year, and in 1937 were $721,871 000 or 21 9 per cent of that year's total Imports of crude materials amounted to $732,965,000 in 1936 and $973 535,000 in 1937, or respectively 30 2 per cent and 32 3 per cent of total imports for consumption in the two years Crude foodstuffs exported in 1936 were valued at $58,144,000 which was 2 4 per cent of total exports for that years, and $101,742 000, or 3 1 per cent of the total in 1937 Imports of crude foodstuffs for consumption were $348,682,000 or 14 4 per cent of the total value of imports for consumption in 1936, and $413,345,000 or 13'7 per cent of the total in 1937. Manufactured foodstuffs exported in 1936 came to $143 798 000 or 5 9 per cent of the year's total and in 1937 were $177,451,000 or 5 4 per cent of the total Imports of manufactured foodstuffs for consumption amounted to *$385,240,000 or 15'9 per cent of the total* imports in 1936 ; and $440,103,000 or 14 6 per cent of the total in 1937 Semi manufactures exported in 1936 were valued at $394,760 000 or 16 3 per cent of the total, in 1937 they were $677 254,000 or 20 6 per cent. of the years exports Imports of semi-manufactures for consumption totalled

$490,238,000 or 20 2 per cent of all imports for consumption in 1936 and $534,181,000 or 21 1 per cent of the total in 1937 Finished manufactures worth $1,154,099,000 of 47 7 per cent of the total for that year were exported in 1936, and $1,616,598 000 worth, or 49 1 per cent of the total, in 1937. Of finished manufactures imported for consumption $465,852,000 worth or 19 2 per cent of all imports for consumption, came in during 1936 and $551 323,000, or 18 3 per cent of the total were received in 1937

(*B A Hons 1944*)

CHAPTER III

MODE, WEIGHTED AVERAGE, GEOMETRIC AND HARMONIC MEANS

Mode

Mode is the predominent item in a series, it is the size of the variable that occurs frequently or the position of the greatest density

Local inquiries into wages frequently require the 'current' wage or the 'usual' wage This wage should be considered as Modal wage.

Inquiries regarding modal wages, rents, price etc, are frequently answered off hand by experienced businessmen whilst enquiries as to Average quantities, would involve a considerable amount of labour Mode is also called Norm

Meteorological forcestes are based on the use of the mode In studying output, Mode proves of great advantage.

To locate the position of mode, the following formulae may be used First Group the data, notice the maximum frequency and apply

(1) $\text{Mode} = l + \dfrac{f_m - f_1}{(f_m - f_1) + (f_m - f_2)} \times i.$

Where l is the lower limit of the modal group, that is the group having the greatest frequency i being its magnitude.

f_m is the maximum frequency, f_1, the frequency of the group preceding the modal group and f_2 of the gro following the modal group

(2) $\text{Mode} = l + \dfrac{f_2}{f_1 + f_2} \times i$

This is more handy than (1) for calculation, but (1) gives more precise position

(3) Mode = 3 **Median**—2 Arithmetic mean.

This formula is quite general for the calculation of mode In case of frequency distribution, where two c more equal maximum frequencies occur, this formula is to be used.

Example 1.—

Marks out of 10	*Number of Students.*
2— 4	20
4— 6	40
5— 8	30
8—10	10

The group (4—6) contains the maximum frequency 40 so it is a modal group with 2 as its magnitude, 40 being th maximum or modal frequency by (1) $\text{Mode} = 4 + \frac{40-20}{20+10} \times 2 = 5\frac{1}{3} = 5\ 33$

(2) $\text{Mode} = 4 + \frac{30}{20+30} \times 2 = 5\frac{1}{5} = 5\ 2$

(3) Mean is $5\frac{3}{5}$, Median $5\frac{1}{2}$

$\therefore \text{Mode} = \frac{33}{2} - \frac{56}{5} = 5\frac{3}{10} = 5\ 3$

Symmetrical distribution —If in a series, the mean, median and the mode are the same, the diftribution is said to be symmetrical otherwise non-symmetrical

Advantages and disadvantages of Mode —

1. Mode is easily understood and like median it may be spotted by inspection, an advantage which the Arithmetic Mean does not enjoy.

2 Like the median and mean, it can be calculated when data fall into groups

3 It is the average of position and proves useful for non-quatitative data also

Disadvantage —It is frequently ill defined and becomes difficult to locate exactly by the formmlæ Its significance is limited when a large number of values is not available.

Weighted Average

The Arithmatic average gives equal importonce to all the items in a series and it cannot be advantageously used where it is necessary to give unequal importance to different items. In such cases weighted average bas to be used

Due importance is given to each item by weighting it.

The object of weighting is to give proper importance to different data Weights are assigned to each item in portion to its importance in influencing the final resul and each item is multiplied by its weight or by the num of persons or things connected with it, and the products added up. The total sum of the products is divided by th sum of weights (or by the number of persons or things nected with it) and the result is the *Weighted Averag* Weighting may be essential when the series is small, very large series, weighted average and Arithmetic averag tend to be the same Weights are estimates of relati importance

Example 2—

Description of workers	FACTORY A.				FACTORY B.		
	No of employees	*Daily wage per employee*			*No of employees*	*Daily wa, per emplo:*	
		Rs	a.	p.		Rs	a.
(*a*)	200	3	8	0	320	2	4
(*b*)	20	1	8	0	40	1	4
(c)	250	2	8	0	300	4	0
(*d*)	150	5	0	0	200	5	0
	620				860		

Simple Arithmetic Average for

$$\text{Factory A} = \text{Rs.}\ \frac{3\cdot5+1\cdot5+2\cdot5+5}{4} = \text{Rs}\ \ 3\cdot125 \ \ \text{which}$$

the same for B also, so no comparison is possible.

Weighted average for Factory B

$$=\frac{(2·25\times320)+(1·25\times40)+(4\times300)+(200\times5)}{320+40+300+200}$$

$=\frac{2977}{860}=3·453$ For factory A, weighted rverage

$$=\frac{3·5\times400+1·5\times20+250\times2·5+5\times150}{200+20+250+150}=3·395$$

us there is a marked difference in average wages

Geometric and Harmonic Meens

Geometric mean is the nth root of the product of n items.

If $a, b, c,$ z are n items then $G=(a.b.c.d\ldots z)^{\frac{1}{n}}$

Thus the G mean of 4 and 9 is $(4\times9)^{\frac{1}{2}}=6$ G. M of 5, 8 and 25 is $(5.8.25)^{\frac{1}{3}}=10$ G M can be easily calculated with the help of logarithms, *i e*, the logarithms (logs) of the items are averaged and the anti logarithm (anti log) of this average will gave the G M The logs can be looked up easily from the Table of Logarithms

Example 3—To find G. mean of (*a*) 20, 5 and 10.

Using log table, $\log G=\frac{1}{3}(\log 20+\log 5+\log 10)$

$=\frac{1}{3}(1·3010+·6990+1·000)=1$

$\therefore G=10$

(*b*) To find G M of the grouped data

	x	f	$\log x$	$f\times\log x$
2—4	3	20	·4771	9·542
4—6	5	40	·6990	27·96
6—8	7	30	·8451	25·353
8—10	9	10	·9542	9·542
		100		72·397

$$\log G=\frac{\Sigma(f\log x)}{\Sigma f}=\frac{72·397}{100}=·72397 \quad \therefore G=5·297$$

Harmonic Mean is the reciprocal of the average of the reciprocals of the items in a series Harmonic mean of n items will be $\dfrac{n}{\frac{1}{a}+\frac{1}{b}+\quad+\frac{1}{z}}$.

Example 4—To find Harmonic Mean for ungrouped as well as grouped data

(a) To find H M of *5* and *25* There are two item and, therefore $n=2$

$$\text{and H M} = \frac{2}{\frac{1}{5}+\frac{1}{25}} = \frac{2}{2+4} = 33$$

(b) x	*Reciprocals* $\frac{1}{x}$	*Frequency* f	$f\times\frac{1}{x}$
3	$\frac{1}{3}=333$	20	6 66
5	$\frac{1}{5}=2$	40	8
7	$\frac{1}{7}=143$	30	4 29
9	$\frac{1}{9}=111$	10	1 11
		100	20 06

$$\text{Harmonic Mean} = \frac{100}{20\ 05} = 4\ 98$$

$$\text{In general H M} = \frac{\Sigma f}{\Sigma\left(f\times\frac{1}{x}\right)}$$

Reciprocals can also be taken from the table

Advantages and disadvages of these Means—Harmonic Mean is less than the Geometric Mean which is less than Arithmetic Mean. If in the data, Arithmetic Mean fails to give a satisfactory average, or the average being too big in comparison with data, then Geometric Mean is to be used and if that also is unsatisfactory, then Harmonic, but Harmonic is not much used in practice.

If two or more series are to be compared and Arithmetic Mean comes out to be the same then Geometric Mean can be used

Geometric Mean is less affected by extremes It is particularly useful in the Construction of Index Numbers

Geometric Mean cannot be determined where there are negative values in the series or where one of the items is zero and moreover it involves lot of calculations

Exercise II

I —Find the Mode for the data in Q 14, Exercise I Are these symmetrical distributions ?

Ans Using formula (I) (a) 6, (b) 43 3, (c) $13\frac{2}{5}$, (d) $5\frac{1}{3}$ (a) is symmetrical.

II —Calculate the G Mean and H M for Q 14, Ex I (*a* and *b*)

Ans (a) 5'615, 5·161, (b) 45 11, 42 15

III —(*a*) Find the Geometric Mean of 50, 80, 200 and 100 and compare with the Arithmetic and Harmonic Mean

Ans 94 57 107½ 84 21.

(*b*) Find the Mode of

0— 4	10
4— 8	20
8—12	30
12—16	30

Ans $10\frac{8}{9}$

IV —Find the Mean, median and Mode of—

Class intervals,	6 5—7 5	7 5—8 5,	8 5—9 5,	9 5—10 5,	10 5—11 5	11 5—12 5,	12 5—13 5
Frequencies	5	12	25	48	32	6	1

Ans 9 87, 9'98 10 2

V —Determine the Mode in Q IV by using formula (2).
Ans 10 06.

VI —Compute the modal wage for the following frequency distribution of wages —

Central wage Rs	15 20, 25, 30, 35, 40, 45, 50, 55
Wage earners	2, 22, 19, 14, 3, 4 6, I 1.

Ans Classify the wages as 12 5—17 5 etc
Apply formula (1) 21 85.

VII – Table showing the frequency with which profits are made What is the Mode ?

						Frequency
Exceeding	Rs	3 000	and not	exceeding	4,000	3
,,	,,	4,000	,,	,,	5,000	7
,,	,	5,000	,,	,,	6,000	22
,,	,,	6 000	,,	,,	7,000	60
,,	,	7,000	,,	,,	8,000	85 ✓
,,	,	8,000	,,	,,	9,000	32
,,	,,	9,000	,,	,,	10,000	9

Ans using (2) Rs 7347 82.

VIII —The annual incomes of fifteen families are given below in Rupees 80, 2,500 90, 1,200, 1,450, 7,200, 120, 1,060, 150, 480, 360, 96, 200, 520, 60 calculate the Arithmetic Average, Geometric Mean and the Harmonic Mean

(P U. M A. 1940). Ans 1037 7, 377 3, 186

IX —The following is the distribution of wages per thousand employs in a certain factory —

Daily wages in annas	1 2 3	3 4 5	5 6 7	8 9	10 11	11 12 13	14	16	18	20	22	-
Number of employees	3	13	43	102	175	220	204	139	69	25	6	1

Calculate the modal and median wages and explain why there is a difference between the two

(E A. (Hons.) 1943) Ans $11\frac{29}{61}, 11\frac{27}{55}$

X.—The following marks have been obtained in three papers of Statistics in an Examination by 12 students. In which paper is the general level of the knowledge of the students highest ? Give reasons,

A 36, 56, 41, 46, 54, 59, 55, 51, 62, 44, 37, 59.

B 58, 54, 21, 51, 59, 46, 65, 31, 68, 41, 70, 36

C 65, 55, 26, 40, 30, 74, 45, 29, 85, 32, 80, 39.

Ans Paper A

XI —Calculate the Average for

Items	*Expenditure*	*Weight*
Food	29	7 5
Rent	54	2
Clothing	97'5	1 5
Fuel and light	75	1
Other items	75	5

Ans. 46 74.

XII—The following table gives the number of employees and their monthly earnings in two factories of a particular city :—

Descrition of workmen	A		B	
	No of employees	*Monthly earnings*	*No. of employees.*	*Monthly earnings*
		Rs.		Rs.
(*a*)	4	800	1	750
(*b*)	22	45	8	125
(*c*)	20	100	10	50
(*d*)	30	30	20	40
(*e*)	80	35	30	45
(*f*)	300	15	100	15

Compare and find the weighted average.

Ans 31'5 and 34 9.

XIII.—Calculate the geometric and harmonic means weights in maunds

250, 12, 4 5, 119·5, 30, 42, 35.4, 75.

Ans 39 8, 19

XIV.—Determine the mode and Geometric Mean : Questions 23—27 (Exercise I) and compare the averages

CHAPTER IV

DIAGRAMS AND GRAPHS

The statistical data can be presented in the form of diagrams, charts, graphs and pictures, so as to permit immediate grasp of the significance attached to · The method of diagrammatic representation is used · the purpose of comparisons In business, it is necessary to call for data relating to Sales, Purchases, Stock Expenses, Cash Balance, etc, and if these are presente to the business man in a graphic form in such a that comparison could be made between two or periods, or two or more related items, it would be easie to understand, than analyse the tabular statements an also save a lot of time. Great care should be taken . the choice of suitable diagrams depicting a concise p of the statistical data The size of the diagram should · just sufficient to enable the eye to perceive the features of the figures which it claims to stand for. Th diagrams should be neatly and accurately drawn with th help of instruments and they should be attractive an complete, as far as possible. To bring out the distinctio clearly, various kinds of dottings, lines, pencils of . .. colours, crossing or colouring or some other · methods may be used

The following types of diagrams, charts and graphs re commonly used

(1) Simple Bar Diagrams, (2) Subdivided Bars ir Compound Bar Diagrams and Percentage Bar Charts, 3) Rectangular Diagrams, (4) Squares Cubes and Circular)iagrams (5) Pictograms, (6) Historigrams, (7) Logirithmic or Ratio Charts (8) Graphs of Frequency)istributions

(1) Bars or thick lines of uniform breadth and with niform space in between, are drawn to represent the iven items, the magnitude being represented along the ertical side of the bar on a convenient scale and the tems arranged in ascending or descending order of agnitude

(2) If a magnitude is capable of being broken into omponent parts or if there are independent quantities hich form the subdivisions of the total, in either of these ases, bars may be subdivided into the ratio of the arious components to show the relationship of the parts) the whole If the imports and exports of a country re given, the sum of the two, the total foreign trade, ill be represented by the height of the bar, imports and xports will be the sub divisions Total population of a ountry may be represented by the height of the bar and e males and females will be the sub divisions (See so Exercise III, 2)

If the sub divisions are more than two, the subvisions may be reduced to percentage of the whole e height of the bar will represent 100 and the other mponents in percentages may be represented on the bar is will be a percentage Bar diagram

(3) Bar diagrams explained above, are supposed to have no breadth at all, but Rectangular diagrams have breadth as well as height

The area of the rectangle will represent a magnitude. Rectangles may be used when two or more quantities are to be compared and each is sub-divided into several components For instance, when it is desired to show differences in expenditure, on the same item, in two family budgets will different incomes, rectangles can be used with incomes as the breadth of the rectangles and 100 as the height of the rectangles. The several items of expenditure may be reduced as percentages and represented on the rectangles. A uniform scale is to be used for the rectangles. *e g* See Exercise III, 6),

(4) *Squares, Cubes and circular diagrams* —When quantities bearing large ratios such as 1 100 or near about, are to be compared, bar diagrams do not serve the purpose as a suitable scale cannot be selected In such cases, squares are used

Take the square roots of the given items (arranged in ascending or descending order) and with these square roots as the sides consruct squares with a convenient scale, keeping a uniform space in between the squares. (e.g., see Exercise III, 7). If the ratios in quantities are 1 : 1000 or near about, *cubes* are drawn with cube roots as sides

As it may take more time to construct squares, circles can be used in place of squares. With square roots of the items [illegible] draw circles of all the items. T centres should be placed in a horizontal line. Circles are also called **Pie Diagrams.**

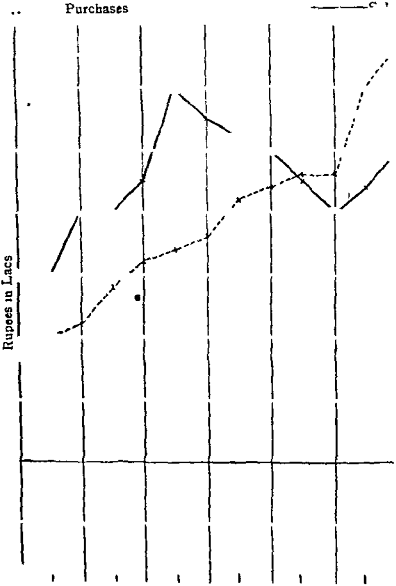
Sectors of the circle can also be used for comparing several items, the sub divisions being represented as follows —

Suppose we are given the population of several countries. Let the circle represent the total of the populations. The whole circle covers 360 degrees, that is (i e) the whole population = 360 Express the populations of other countries in degrees and draw these angles in the circle. The sectors so formed will represent the different populations If the total population is 120 millions and of one country is 10 millions, then the angle of the country is $\frac{360}{120}\times10=30$. The sector countaining 30° will represent the population of the country

(5) *Pictograms* —Numerical data are generally given a beautiful and attractive summary representation by means of appropriate maps or cartograms, and pictures or pictograms In drawing pictures it should be borne in mind that the proportions in wihch the natural objects are found should not be disturbed. Maps with different colours may be used to visualise the distribution of population in an impressive manner For ocular comparison of figures dotted maps are widely used. Density of population, the average yield per acre of corps in various parts of a country and many other similar statistics may be indicated by means of dots in a map.

(6) *Historigrams* —Diagrams pertaining to historical or time series are said to be historigrams The years

HISTORIGRAMS



or months as the given time may be are plotted along the horizontal line (called the axis of x) on the graph paper the data corresponding to the periods are plotted along the vertical line (called the axis of y) This plotted points are joined by straight lines This will be a Historigram drawn on a convenient scale The point where the axis of x and the axis of y meet is called the origin which is zero for vertical values If there are two or more series on the same periods they can be plotted on a convenient scale with origin as zero for vertical values and thus their fluctuations can be compared Commercial data such as Records of sales Purchases and Sales Gross Profits and Expenses Turnover and Net Profit can thus be represented graphically Some distinction may be made when there are two or more time series

Example—To draw the Historgrams for the data showing purchase and sale for 12 months given in Lacs of Rupees

	January	Feb	March	April	May	June
Purchases	40	42	48	52	54	56
Sale	50	60	60	65	80	74

	July	August	Sept	Oct	Nov	Decembe
Purchases	61	64	66	66	80	86
Sales	72	70	65	60	64	70

The months are shown along the horizontal axis and the rupees in lacs along the Vertical axis in the diagram

(7) *Logarithmic or Ratio graphs*—So far we have been dealing with data drawn on the natural scale th

, equal vertical distances represent equal absolute movements The ratio scale is employed as an alternative , the natural scale, whenever it is desired to study lative movements An absolute series may be 'converted into a ratio series by plotting either (1) the ogarithms of the actual figures of the given items, p be represented along the Vertical axis or (2) the gures themselves on a semi-logarithmic paper. Method) is generally used as the logarithmic paper is not isily available The logarithms can be looked from the ables and then plotted The plotted points may be ined by means of straight lines to obtain a Logithmic Graph (or Ratio Chart) or by a free hand curve hen possible Ratio scale cannot show zero and negave values which the natural scale can A constant rate change, growth or decline is indicated by a straight ne on a logarithmic graph The stability or instability of ices or any other such variable can be brought out by the garithmic graph

Example —To draw a population graph from the fol wing data on a ratio scale for the population of India lacs

ears	1881,	1891,	1901,	1911,	1921,	1931
opulation	2539,	2873,	2944,	3150,	3189,	3530

The logarithms of the figures in population, are, 8·4048, 8·4584, 8·4689, 8·4983 8·5037, 8·5478

lotting these as in historigrams, we get the required graph pproximate value in decimals may be taken while lotting).

example given above, we are given the population during the years 1881—1931 If we are required to estimate the population for any intervening year say 1926, not given in the data, interpolation has to be used as follows Mark the year (say 1926) along the axis of x, and at this point erect a perpendicular (called Ordinate) cutting the graph at a certain point The length of this Ordinate will indicate an estimate of the population for 1926 Looking its value from Logarithmic tables, we shall have the estimate of the population. Extrapolation can also be done graphically if the data happen to be organic in character It means, finding the value for the year beyond the years given in the data, *i.e*, after 1931 in this example Plot the year (say 1941) along the x axis and erect a perpendicular Extend the drawn graph carefully in continuation with its trend beyond 1931, and let it cut the perpendicular at a certain point. The length of this ordinate after consulting the log-table will give the estimate of the population of 1941

Extrapolation or forecasting will depend upon the constant rate of increase of the graph and on economic and other conditions governing the data

For interpolation, in general plot the observations along the x axis and y-axis Join the points by a freehand curve. To find a value of y corresponding to any value of x, erect a perpendicular through that point on x axis cutting the curve at a certain point. Read the value of this ordinate This will be an estimate of the interpolated value In time series the missing values for any particular year can thus approximately be found

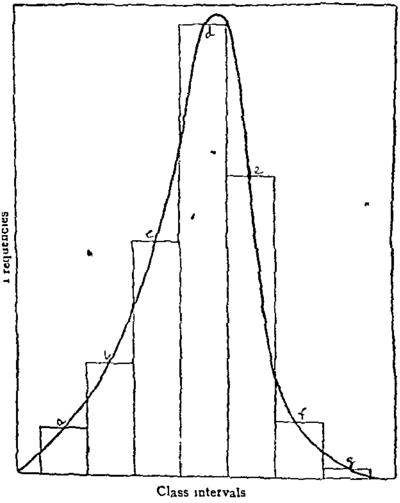
(8) *Graphs of Frequency Distributions*—Frequency distributions *are represented graphically by (i) Histograms* or Column diagrams or Block diagrams, (ii) Frequency polygons and frequency curve, (iii) Cumulative frequency curve.

(i) *Histogram*—Plot the class intervals along the axis of x, side by side Take the first class interval and draw on it a rectangle with the corresponding frequency as height Take the second class interval and draw a rectangle with its corresponding frequency as height along with the first rectangle In this way draw all the rectangles along with each other for the whole frequency distribution The set *of rectangles so drawn will form a Histogram* This is in general use for a frequency distribution

(ii) *Frequency polygon and curve*—Mark the central values of the class intervals along x axis, and plot the frequencies corresponding to the central values Join the plotted points by means of straight lines, the figure so formed will be a frequency polygon In a histogram if the *middle points of the top horizontal sides of the rectangles* are joined by straight lines the figure formed is a frequency polygon and if these middle points are joined by means of a smooth freehand curve, the curve formed is a frequency curve or smoothed histogram This curve in most of the cases is bell shaped in form and is such that its area is approximately the same as that of the polygon or the rectangles The diagram drawn on the annexed page is that of a Histogram and frequency curve The middle points a, b, c, d e f when joined by lines will give a frequency polygon

(iii) *Cumulative frequency curve or ogive*—Form the cumulative frequencies and plot these along the vertical line

Histogram and Frequency Curve



at the upper limits of the class intervals, marked along the horizontal axis Join the plotted points by means of a freehand curve. The curve drawn will be a cumulative frequency curve or ogive

For joining the points lines can also be drawn if the series is discrete but for continuous series where class intervals are small and number of observations great, freehand curve should be drawn

The ogive is useful for locating graphically the Median, and Quartiles as follows Along the vertical axis mark the total number of frequencies and also its middle point for median From this middle point draw a line parallel to the x axis, cutting the ogive at certain point The distance o this point from the vertical will give the value of the Median according to the scale used In the same way, to obtain first and hird Quartiles mark ¼th and ¾th distances instead of the middle point and proceed as for the median

Interpolation may also be carried along the ogive Deciles and Percentiles can also be located

Example—To draw the cumulative frequency curve for the following frequency distribution and locate the Median

Class intervals	*Frequencies*	*Cumulative Frequencies*
1—5	4	4
6—10	10	14
11—15	28	42
16—20	49	91
21—25	58	149
26—30	82	231
31—35	87	318
36—40	79	397
41—45	50	447
46—50	37	484
51—55	22	506

The adjoining diagram gives tne cumulative frequency curve The cumulatives areplotted against the upper limits of the class intervals 5 10 15   55 The points are joired by means of a freehand curve

To locate the median mark the total frequencies 506 along the vertical OY and then its middle po nt From this middle point draw a line pa allel to the horizontal OX' cutting the curve at the point The distance of th s point from the vertical will give the med an value

Percentage cumulative frequency curve —To draw this curve first express each frequency as a percentage of the total number then form the cumulatives and plot them as in the case of the ogive The curve drawn will be a cumulative percentage frequency cu ve and the able formed will be a percentage frequency table

The curve is useful for comparing and adjusting the distributions

Histogram for unequal interval —If a frequencry distribution consists of same equal class intervals and a few unequal c ass intervals the histogram can be drawn as follows

Mark the class intervals along the *x* axis and erect rectangles on the class intervals of equal magnitude with their corresponding frequencies as ordinates For unequal class intervals notice the relation of their magnitudes to the equal class intervals If the unequal magnitudes are *m* times say) the equal magnitude then divide the frequencies orresponding to the unequal class intervals by *m* and aking these as Ordinates draw rectangles on the the inequal class intervals The set forectangles so formed will be a H stogram (e g Exercise III 15)

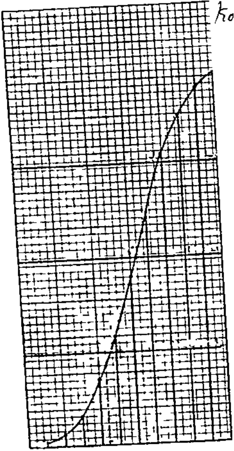
Lorenz Curve and Pareto Curve

Lorenz curve is employed in order to measure the oncentration of wealth or income This curve takes the orm of a cumulative percentage frequency curve combining he percentage of items under review with the percentage of wealth or other factor d stributed among such items

S/ Nemi Chand

Cumulative Frequency Curve

Kota



It is useful for comparing the distribution of prc over differerent groups of business and showing d of a group The following example illustrates the methoo of the construction of the Lorenz Curve Consider follow ng data relating to the distribution of estates exceedi £10 000 in net capital Value

Number and Capital Values of Estates in Great Brit liable to Esta e Duty 1929 30

1 *Capital value exceeding* (1) (£000)	*Cumulative Number of Estates* (2)	*Cumulative Percentage* (3)	*Cumulative Net Capital Values* (4) (£000 000)	*Cclumn (as Percentages* (5)
3 000	2	0 02	12 4	3 39
2,000	6	0 07	16 2	4 42
1,500	10	0 11	24 7	6 74
1 000	15	0 17	32 7	8 93
800	20	0 23	36	9 83
600	35	0 40	47 1	12 86
500	45	0 55	52 6	14 36
400	68	0 78	60 1	16 41
300	119	1 37	77 5	21 16
250	158	1 81	86 4	23 59
200	214	2 46	100	27 31
150	317	3 64	118 4	32 33
100	581	6 67	149 5	40 82
80	817	9 38	169 7	46 34
60	1172	13 46	195 2	53 3
50	1467	16 84	211 7	57 81
40	1971	22 63	233 8	63 84
30	2804	32 19	262 3	71 63
25	3420	39 26	279 8	76 41
20	4418	50 72	302 7	82 66
15	5923	68 00	329 6	90 00
10	8710		366 2	100 00

1 Connor Statistics / in theory and Practice pages 203 4

Procedure —(*a*) Convert the cumulatives in column (2) in percentages, total being 8710 if in column (2) cumulatives are not given then first take cumulattive and then percentages or first form the percentages and then take the cumulatives These are given in column (3)

(*b*) Convert column (4) into percentages, total being 366 2 If cumulatives are not given in any distribution take the cumulatives and then percentages.

Draw the graph of the cumulative percentages in columns (3) and (5) The curve traced will be the Lorenz Curve, as shown in the diagram

The straight line joining the extremities denotes the line of equal or even distribution. The concavity of the curve away from the straight line is a measure of con tration of wealth

By drawing two or more Lorenz curves, we may compare income distributions at different times or places

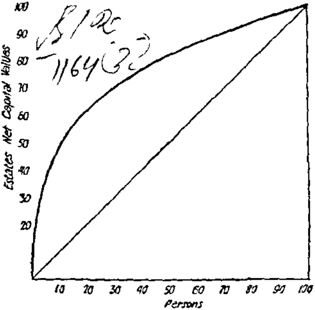
Pareto s Law —If a cumulative frequency distribution of incomes be plotted upon a double logarithmic scale, the points will lie approximately upon a straight line

This is Pareto s law after Pareto (Italian) This statement is true of Great Britain, United States, Germany, British India and other countries where it has been tested

The following table and graph illustrates the Pareto s Law, with reference to Great Britain and Northern Ireland.

Cumulative distribution of Incomes 1928 29.

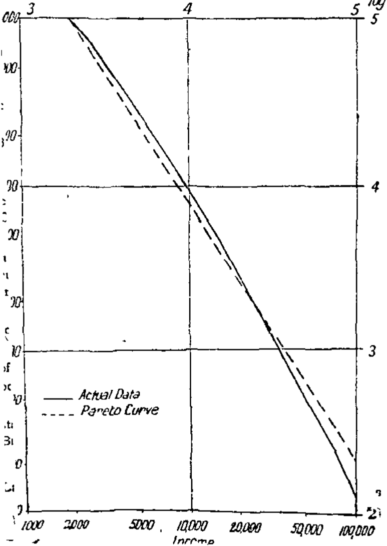
Lorenz Curve



S/25
1164(3)

Pare to Curve

Year 1928-29



Income (x) (1)	Number of Incomes of £sx, of over (y) (2)	Log (x) (3)	Log (y) (4)
2,000	97 696	3 3010	4 9899
2,500	7+,211	3 3979	4 8705
3,000	57 878	3 4771	4 7625
4,000	38 539	3 6021	4 5859
5 000	27 722	3 6990	4 4428
6 0 0	20 975	3 7782	4 3217
7,000	16,544	3 8451	4 2186
8,000	13,317	3 9031	4 124+
10 000	9,163	4 0000	3 9620
15,0 0	4,595	4 1761	3 6623
20 000	2 781	4 3010	3 4442
25,000	1 851	4 570	3 2674
30,000	1 324	4 4771	3 1219
40,000	753	4 6021	2 8768
50 000	487	4 699	2 6875
75,000	234	4 8751	2 3692
1,00,000	130	5 0000	2 1139

1 Connor, page 200—203

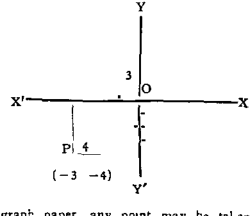
Column (1) shows the income (x) and column (2) the b number of incomes of £ x or over Columns (3, and (4 show the logarithms of the figures in columns (1) and (2)

Plotting the ogrithms we get Pareto s curve, whict is approximately a straight line The steeper the slop, of the curve, the more equally is income distributed and *vic versa*.

Pareto s law is not ecognised as a general law of in v come distribution Pareto s curve can be used for interpolatioi n and not for extrapolation Mathematically the law i w $y = ax^{-b}$, where y is the number of persons whose incom se is at least x units (Rupees pound £s etc) a and b are cor stants which depend on the country or the class of the community that is being considered In logarithms, the equation can be written as

$\log y = \log a - b \log x$ b is the slope of the curve, its usual value being 1·5 nearly

Curves of the type $y=ax^b$ and $y=ab^x$



On a graph paper, any point may be taken as origin where the value of the variables x and y will be zero The positive values of x and y are measured along the lines OX and OY respectively The negative values of x and y are measured along OX′ and OY′ respectively XOY s the first Quadrant having x and y positive

A point is represented by the values of x and y and s written as (x, y)

In the first Quadrant XOY x and y are positives

In the second Quadrant YOX′, x is negative (−ve) but y ositive (+ve)

In the third Q uadrant X'OY', x is $-$ve and y is $-$ the point (pt) will be $(-x, -y)$

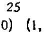
In the fourth Quadrant XOY', x is +ve and y is $-$ the pt. being $(x, -y)$ The value of x and y, $x=-3$ a $y=-4$ are called the co-ordinates of a point say P plotted in the fourth quadrant Such a graphical system known as Cartesian system.

In the curve $y=ax^{b}$, a and b are constants, but x y are variable and can have any values In the equati $y=ax^{b}$, y depends upon x, so x is called the ind variable and y the dependant variable Let us consider tl well known cases of this general equation when b is positi say $y=x$, $y=x^2$, $y=x^3$

Allot values to x, as shown,
To trace $y=x^2$, the values can be allotted as

x	0,	1	2,	3	4,	5 ,	-1,	-2,	-3,	-4,	-5 .
y	0	1,	4,	9,	1o,	25 ,	1,	4,	9,	16,	25

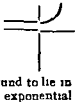
Plotting these points with co ordinate (0, 0) (1, (2, 4) – we get the required graph , both the branch going at an indefinite distance or infinity, such a curve called the Parabola and the equation represents a Parabolic curve lying in the first and second quadrants. If the equation had been $y^2=x$ the parabola will lie in the first and fourth Quadrants.



The graphs of curves $y=x^4$, $y=x^6$, $y={}^8$ i e., of ev powers of b, will all lie in the first and second Quadran The graphs of odd powers of x, i.e., of $y=x^3$, $y=x^5$—w lie in the first and third Quadrants, and $y=x$ will represe a straight line

The above method of plotting curves is general and is ipplied for all curves in Cartesian system.

Exponential Curves —The curves given by, $y=ab^x$ are alled Exponential curves The curve is drawn by plotting he points, as shown in $y=4^x$



Points are x 0, $\frac{1}{2}$, 1, 2, 3,
y 1, 2 4, 16, 64,
$-\frac{1}{2}$, -1,
$\frac{1}{2}=5$, $\frac{1}{4}=25$.

Plotting out these values, the curve is found to lie in he first and second Quadrant In this way exponential urves of the form, $y=ab^x$ can be drawn

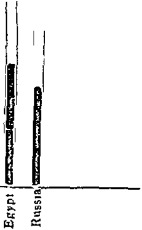
Exercise III

1. Draw a bar Diagram to represent the turnover of a ompany for 12 years

Rs. 35,000, 42,000, 43,500, 48,000, 48,500, 52,000, 36,500, 54,500, 100,000, 54,000, 112,500, 194 000

2 The following table gives the Birth Rates and)eath Rates per thousand of a few countries Represent 1em by a Diagram (Sub divided)

Country	Birth Rate	Death Rate
ndia	33	24
apan	32	19
ermany	16	10
gypt	44	24
ustralia	20	9
lewzeland	18	8
rance	21	16
ussia	38	16



3, Draw a percentage bar diagram for Birth Rates Death Rates in Q. 2.

4 Represent the following figures about infant m tality in different cities by a suitable diagram,

London	Calcutta	Bombay	Naghpur	Madras	Par
66	244	274	323	251	93

5 Draw the graph of the following time series —

Years	*Gross Profit* *Rs.*	*Expenses* *Rs.*	*Net P.* *Rs*
1	7,900	2,700	5,200
2	5,500	1,700	3,800
3	4,800	1,500	3,300
4	4,500	1,000	3,500
5	6,500	4,000	2 500
6	9,000	5.000	4,000
7	8,500	3,500	5,000
8	7,000	3,000	4,000
9	6,500	1,600	4,900
10	6,200	2,500	3,700

6 The following table gives details of the month expenditure of three families Represent them by a suitab diagram

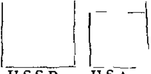
Items of Expenditure	.	Family A		Family B	Famil C
		Rs	a.	Rs.	Rs.
Food	..	12	0	25	30
Clothing	...	2	8	8	10
House rent		2	0	4	8
Education	.	1	0	5	7
Miscellaneous	.	2	8	8	15
Total	..	20	0	50	70

This is an example of Ractangular Diagram.

7. Draw squares for the following table which giv

ɔe production of wheat of the following countries ın a cer ıın year.

'ountries	Quintals (000,000)
ınited Kingdom	12
ɔdia	105
gypt	11
S A.	230
Africa	3
anada	108
.S S. R	289



8 Draw the following diagrams on a logarithmic scale ɔr Q 8 and 9.

United Ringdom Receipts for Super Tax

Year.	1911,	1912,	ı913	1914,	1915,	1916,
£(000)	2891,	3018,	3600,	3339,	10120	16788,
	1917,	1918,	1919,	1920,	1921,	
	19140,	23280,	35560,	42405	55669,	
	1922,	1923,	1924,	1925,	1926	
	61350	63910,	61747,	62989,	67835	

9. *Acreage of crops*

	(A) (000) acres	(B) acres
1920	1949	13050
1921	2040	8335
1922	2034	8415
1923	1799	16923
1924	1594	32637
1925	1550	56243

10 The following figures gives the quantity of sugar ᴾoduction ın the followirg countries Represent them (1) ɣ circles (2) by sectors (3) by cubes

	Production of Sugar in Quintals (000,000)
India	20
Egypt	1
Cuba	32
Java	30
Australia	5
Japan	1

11. Represent the data in Q 15 (Ex I) by a suitable diagram

12 Draw Histogram, frequency polygon and curve for the data in Q. 16, 11 and 12 (Ex. I).

13. Draw the frequency graph for Q. 23 (Ex. I)

14 Draw the cumulative frequency curve for Q. 26 27 and 24 (Ex I) and locate the median and Quartiles Compare the values obtained by actual calculation.

15 Data showing the Intelligence Ratios of 1000 children Draw a Histogram.

125—139	23	85—89	139
115—124	40	80—84	133
110—119	37	75—79	89
105—109	71	65—74	83
100—104	90	45—64	29
95— 90	134		
90— 94	132		1000

Hint.—This is an example in which class intervals are written from the highest to the lowest. It is to be noticed that the last two class intervals are of unequal magnitude. Therefore the frequencies, for the purpose of drawing will be ; for 45—64, which will have a base four times, than the equal interval (5) the frequency $\frac{1}{4}\times 29$, in order that its area nay represent a frequency of 29 Similarly the height of he Ra tangle representing the frequency of 65−74 will be $\frac{1}{4}\times 83$ Rectangles are to be drawn side by side

16 Represent the following graphically

	(A)	(B)	Total
1935	130	370	500
1936	110	260	370
1937	134	256	390
1938	146	244	390
1939	159	285	445

Hint —This is a subdivided Bar diagram for each year with the given totals

17 Draw a Histogram and frequency pulygan for the following distribution

Degrees of cloudiness (x)	*Frequency*	*Degrees of cloudiness* (x)	*Frequency*
10	580	4	45
9	150	3	68
8	196	2	75
7	75	1	130
6	55	0	220
5	40		

Hint —Take x as the central values and then plot.

18 Draw a cumulative percentage frequency curve for Q 15

19 Draw the graphs of the following curves $y=x$,

$y=x^3$ $y=x^4$, $y=\frac{1}{x}$, $x^2=y$ $y=2^x$, $y=\frac{1}{x^2}$

20 Draw Lorenz curves for the comparison of profi of two groups A and B in business

Total amount of profits earned by Companies in each Division	*No of Companies in each Division*	
Rs	*Group A*	*Group B*
600	6	1
2500	11	19
6000	13	26
8400	14	14
10 500	15	14
15 000	17	13
17 000	10	6
40 000	14	

21 Construct an ogive curve for the following frequency d stribution of Cotton Mills in Bombay according t the quanity of Cotton consumed and estimate the value o the median from the curve

Cotton Consumed in thousand candies	*No of Mills*	*Cotton Consumed in thousand candies*	*No of Mills*
0— 2	5	10— 12	4
2— 4	13	12— 14	1
4— 6	12	14— 16	3
6— 8	11	16— 18	1
8—10	8	18— 20	1
		over 20	2

(B A *Hons 1941*)

22 Represent diagramatically the following data regarding the operation of irrigation works in India

Province		*Area irrigated in Acres*
		Rabi (1926 27)
Madras	—	1,003 065
Bombay		1,128,594
Bengal		447
U P		1,778 645
Punjab		6 084,838
Bihar & Orissa		97,858
C P & Berar		9,165
N W F Province	—	182 574
Baluchistan	—	11 470
Ajmer Mewar	—	22 550

(*B* A *Hons 1941*)

23 The following table gives the populatoin of the United Kingdom and India at the time of the last seven ensuses —

Years		*Population in lacs* *United* *Kingdom*	*India*
1871		315	2062
1881		349	2539
1891		377	2873
1901		415	2944
1911		452	3152
1921	—	471	3189
1931	—	490	3515

Represent the above figures by curves in a logarithmic cale Fstimate the population for 1941

(*M A 1939*)

24 From the date given in Q 17 Exercise I, draw the graph of the accumulated frequencies and hence obtain the value of the median

(Indian Audit and Accounts Service Exam 1941.)

25 Draw a commulative frequency graph of the distribution given in Q 18, Exercise I, and calculate the values of the median and Quartiles (*M A 1943*)

26 Draw a Bar' or 'Pie Diagram to represent the following data —

Output and cost of Production of Coal

Cost per ton disposable commercially	*1924*	*1928*
Wages –	12 74	7 95
Other costs	5 46	4 51
Royalties –	0 54	0 50
Total	18 76	12 96
Proceeds of Sale per ton	19 91	12 16
Profit (+) or loss (−) per ton	1 15	—0 80

(B A Hons 1943)

27 The following frequency distributions shows the number of live stock held by 100 farmers in a tahsil of Bombay Province Draw a graph showing the comulative frequency curve for this distribution and find the two Quartiles and the median (*M. A 1942*)

Live stock units 1, 2, 3, 4, 5, 6, 7

Number of farmers 1, 13, 30, 25, 16, 9, 6 = 100

(M A 1942)

28 Represent graphically the following data for capital outlay and Gross earnings of class I railways in India —

Years	*(In Millions of pounds)* *Capital outlay*	 *Gross earnings*
1923 24	464	70
1924 25	473	74
1925 26	487	73
1926 27	505	72
1927-28	594	86
1928-29	599	86
1929 30	617	84
1930 31	627	77
1931 32	631	71
1932-33	638	70
1933-34	635	72

(B.A *Hons 1942*)

29 The following 44 figures give in arbitrary units the measurements of hardness on different specimens of a certain aluminium die casting —

Specimen	*Hardness*	*Specimen*	*Hardness*
1	53 0	23	64 3
2	70 2	24	82 7
3	84 3	25	55 7
4	55 3	26	70 5
5	78 5	27	87 5
6	63 5	28	50 7
7	71 4	29	72 3
8	53 4	30	49 5
o	82 5	31	71 5
10	67 3	32	52 7
11	69 5	33	7 56
12	73	34	63 7
13	55 7	35	69 2
14	85 8	36	61 4
15	95 4	37	83 7
16	51 1	38	94 7
17	74 4	39	70 2
18	54 1	40	80 4
19	77 8	41	76 7
20	52 4	42	82 9
21	69 1	43	55 0
22	53 5	44	84 8

Group the data into a frequency distribution and draw the corresponding histogram and frequency polygon

(*B A Hons 1942*)

30 The following table gives the number of motor cars produced in three countries during the years 1929—1937 —

(Figures are given in thousands)

Year	Germany	France.	United Kingdom
1929	96	254	241
1930	74	231	241
1931	68	201	226
1932	50	172	246
1933	99	189	296
1934	172	187	355
1935	245	166	417
1936	302	203	481
1937	332	200	493

Represent the above figures by curves on the same graph paper and give necessary comments.

(*M A 1941*)

31 The following table gives the birth rate and death rate of a few countries of the world during the year 1937, —

Name of country	*Birth Rate*	*Death Rate*
Egypt —	43 5	27 2
Canada	19 8	10 2
United States	17 0	11 2
Mexico	40 0	23 9
Argentine	24 0	11 9
India	34 5	22 4
Japan	30 6	17 0
Germany	18 8	11 7
Austria	12 8	13 4
France —	14 7	15 0
Norway	15 3	10 4
England and Wales	14 9	12 4
Switzerland	15 0	11 3
Australia —	17 4	9 4

Represent the above figures by a suitable diagram

(*M A 1941*)

CHAPTER V

DISPERSION OR VARIABILITY AND SKEWNESS

The Average or the typical value is not of much us unless the degree of Variation which occurs about it is in other words, it should be known as to what extent the average is typical, or how the items vary in size

Dispersion or Scatter or Variation or Variability is Measure of the extent to which the individual items vary I the scatter about the measure of central tendency is large, it is of little use as a typical value

Measures of Dispersionare also called Averages of the second order

Measures of Disperson are

(1) The Range, (2) Quartile Deviation or Semi interquartile Range, (3) Mean Deviation or Average Deviation, (4) Standard Deviation,

The Range, the simplest of the Measures, is the difference between the minimum and maximum (smallest and the largest) items in a series As the range depends upon size of extreme items, it is not a satisfactory measure of Dispersion

In the series 60, 61, 63, 65, 67, 68, 90

Range is $90-60=30$

(2) Quartile Deviation or Semi interquartile range is given by $\frac{Q_3-Q_1}{2}$, where Q_1 and Q_3 are the lower and upper quartiles

(3) Mean Deviation is generally calculated from the median. It can also be calculated from the Arithmetic Mean It is the average of the deviations of the items from the Median or Mean deviations being taken positively or Mean Deviation $= \frac{\Sigma |d|}{n}$ where d stands for deviation from Median or (Mean) *taken positively* neglecting negative signs, n being the number of items in the eries For grouped data, Mean Deviation $= \frac{\Sigma f|d|}{n}$ where d stands for deviation of the central values from the Median or Mean n being the sum of frequencies

$|d|$ indicates deviations taken positively

Example 1.

Class intervals	*Central values*	*d*	*frequencies*	*f d*
2—4	3	−2	3	6
4—6	5	0	4 7	0
6—8	7	2	2	4
8—10	9	4	1	4
			10	14

Median $= 4 + \frac{2}{4}(\frac{10}{2} - 3) = 5$ and

$$M\ D = \frac{\Sigma f\ d|}{n} = \frac{14}{10} - 1\ 4$$

In a frequency distribution with unequal class intervals the Arith Mean instead of the Median should be used

Standard Deviation and Variance —Standard Deviation is calculated from the Arith Mean It is given by the formula (1) for ungrouped data

$s\ d$ or σ or $S = \sqrt{\frac{\Sigma a^2}{n}}$, where d stands for deviations of the items from the Arithmetic Mean n being the nu of items (2) for grouped data

σ or $S = \sqrt{\frac{\Sigma f(d)^2}{n}}$ where d stands for the deviation of the central values from the Arithmetic Mean n being th sum of all the frequences $= \Sigma f$ The square of the standar deviation is called Variance

Ex 2—To find σ for 1 2 3, 4 and 5 Arithmetic Mean $= \frac{15}{5} = 3$ Squares of the deviations of these items from 3 4 1 0, 1 4

$$\sigma^2 = \frac{4+1+0+1+4}{5} = 2 \qquad S\ d \text{ or } \sigma = \sqrt{2} = 1\ 414$$

Ex 3—To find the Variance and standard deviation the following freqnency distribution —

	x	f		d	d^2	fd^2
1—3	2	40	80	−2	4	160
3 5	4	30	120	0	0	0
5—7	6	20	120	2	4	80
—9	8	10	80	4	16	160
		100				400

Here Arithmetic Mean $= \frac{\Sigma f x}{\Sigma f} = \frac{400}{100} = 4$

$\Sigma f d = 400^0$ and $n = \Sigma f = 100$

Variance $\sigma^2 = \frac{400}{100} = 4$ and

Standard deviation $\sigma = 2$

Short cut method for finding the Standard Deviation

The short cut method avoids the labour of finding the Aritl metic Mean. Any convenient Provisional Mean can be take and the following formula is then used

$\sigma = \sqrt{\frac{\Sigma f(D^2)}{n} - \left(\frac{\Sigma f D}{n}\right)^2}$ where D stands for deviations of the central values from the Provisional Mean

In the case of the ungrouped data, the above formula is used without f D being the deviations of the items, from Provisional Mean n being the total number of items *i.e.*

$$\sigma^2 = \frac{\Sigma D^2}{n} - \left(\frac{\Sigma D}{n}\right)^2$$

Example 3 can be solved by taking a Provisional Mean ay 6 thus

			$f\,D$	D^2	$f(D)^2$
x	f	D	-160	16	640
2	40	-4	$-\ 60$	4	120
4	30	-2	0	0	0
6 ← o	20	0	20	4	40
8	10 100	2	-200		800

Now $\sigma^2 = \frac{800}{100} - \left(-\frac{200}{100}\right)^2 = 8 - 4 = 4$

therefore ( ) $\sigma = 2$ as before

Characteristics of standard deviation

The *s d* is affected by the value of each item It is the best measure of dispersion It is the least erratic, is suitable for arithmetic and algebraic manipulation and is used for higher statistical operations while the Mean Deviation is not further used

Quartile Deviation is easier to calculate than standard deviation, but it is liable to be erratic.

Relative Measures of Dispersion

The measures given above are absolute measures of dispersion and the resulting values cannot always be compared with significance

To relate the measure of dispersion to its average to convert it to percentage form the standard deviation divided by Arithmetic Mean This measure is known as **Coefficient of Variation** given by $CV = \frac{100\sigma}{\text{Mean}}$ and is generally used for comparison of Variations or Varibility of two or more quantities $\frac{\sigma}{\text{Mean}}$ is called the co efficient Standard Deviation In Example 3 $CV = \frac{100\times 2}{4} = 50$

Other comparative co efficients of dispersion are

Quartile co efficient of Dispersion $= \frac{Q_3 - Q_1}{Q_3 + Q} \times 100$

Mean co efficient of Dispersion

$$= \frac{\text{Mean Deviation} \times 100}{\text{Median or Arithmetic Mean if used)}}$$

✓ **Skewness** —Besides Average and Dispersion skewness is also a measure to study the distributions Skewness is a term for the degree of distortion from symmetry When a distribution is symmetrical the values of the Mean Median and Mode coincide Skewness has the effect of pulling the median and Mean away from the Mode sometimes to the right and sometimes to the left When the Mean is greater than the Mode Skewness is said to be positive it is negative when Mean is less than the mode

A large number of frequency distributions occurring in practice, fall into four types —the symmetrical, the moderately skewed or asymmetrical, the extremely skewed or J-shaped, (in the form of alphabet J), and the U-shaped type (in the form of the alphabet U).

The figure for symmetrical curve will be found in Normal Curve. (See Chap. XI) The somewhat departure for this shape will give a moderatly skewed curve

The co-efficient of skewness that is the measure of skewness commonly used is given by the formula

$$(1)\ S_k = \frac{\text{Mean} - \text{Mode}}{\sigma} \text{ or } \frac{3(\text{Mean} - \text{Median})}{\sigma}$$

For a symmetrical distribution, the co efficient of skewness will be zero.

The second formula is based upon the fact that in a skewed distribution the median does not lie exactly half way between the Quartiles

The co-efficent is also given by

$$\frac{Q_3 + Q_1 - 2\ \text{Median}}{Q_3 - Q_1}$$

The two methods are based on entirely different principles and the results obtained will be different

For a symmetrical and moderately skewed distributions, mean deviation is about $\frac{4}{5}$ standard deviation and the Quartile Deviation is $\frac{2}{3}\sigma$ (approximately).

Exercise IV

I —Find the mean deviation and mean co efficient of Dispersion for

Weekly	*Wages*	*Workers*
Rs	2 4	20
,,	4 6	40
,,	6-8	30
,,	8-10	10

Ans $\frac{3}{2}$;·27

—

II.—Find the average deviation and the standard d. viation of the following — n=8

(a) Rs 300, 400, 700, 200, 600, 500, 100.

(b) Rs 120, 60, 80, 20, 100, 40, 140.

Ans (a) *Rs* 171 4, σ=?

(b) 34 28, σ=

III—Find the standard deviation of the height 10 men Inches 64, 65, 73, 70, 70, 70, 69, 68, 66, 75

Ans 3 15

IV.—Calculate the Mean deviation (M D) from the Median and the Mean, and compare with the stand deviation.

Rs 20, 18, 16 14, 12 10, 8, 6.

Frequencies 2, 4, 9, 18, 27, 25, 14, 1.

Ans *Median*, A Mean *M.D*, *A Mean*, Median *M D*, σ

11 74, 2·29 12, 2 24 2 11

V—Compute the S D and Q D. co efficient of variation and of skewness for the frequency distribution of wages

Monthly Wages		*No of Wage earners*
Rs.	12 5—17·5	2
,,	17·5 22 5	22
,,	22 5—27·5	19
,,	22 5—32 5	14
,,	32 5—37 5	3
,,	37 5—42 5	4
,,	42 5—47 5	6
,,	47·5—52·5	1
,,	52 5—57 5	1
		72

(M Sc, Agri. 1943
Punjab University

$\sigma=8·85$ ✓
$CV=31·8$, ✓
$QD=5·145$, ✓
$SK=·7$ nearly

VI —Calculate the mean and standard deviation of the following values of the World's annual gold output (in millions of pounds) for 20 different years —

94 95 96, 93, 87, 79 73 69, 68, 67, 78 82 83, 89, 95, 103, 108, 117, 130, 97.

Also calculate the percetage of cases lying outside the mean at distances $\pm S$, $\pm 2S$, $\pm 3S$, where S denotes standared deviation

(B A Hons 1942).

D^2 *Ans Mean, 90·15 S = 15 59 35%, 5, and 0.*

VII—From the following frequency table of Marks obtained in Practical Exam Calculate the co-efficient of skewness.

Marks	10	11	12	13	14	15	16	17	18
Candidates	26,	201,	673,	1001,	739,	310	80	13	1

(Aligarh M.A., 1938) Ans 108

VIII.—In Q. 18 Ex 1 find the standard deviation and the mean Deviation

(M A 1943) Ans 543, ·432

IX —Calculate the standard deviation of the chest measurements in Q. 19, Ex I *Ans. 2·05.*

(Punjab M A. 1943. Aligarh M.A. 1941.)

X.—Obtain the standard deviation for the distribution given in Q. 17, Ex. I.

(Indian Audit and Accts Exam 1941) Ans 5 52.

XI.—Compute the standard deviation of the rainfall in — — districts of Bengal from the following

District — 24-Parganas, Murshidabad,
Rainfall in inches — 17·36 — 19·17
(1939 July)

Khul, Burdwan, Midnapur,
22·99, 17 14·19

Rajshai Dacca, Chittagong,
21·23 27·10 40 97

Cooch-Bihar, Hoogly
26·58 17 67

(*B.A Hons. 1941*) *Ans 7 356*

XII —Calculate the co-efficient of variation for the production of Motor-cars. Germany, France and United Kingdom, data given in Q 30 Exercise III

(*M.A 1941.*) *Ans 63 59 , 13 01 , 30 35.*

XIII —Data for Weekly Records of Temperature (Fahrenheit).

Temperature limits — 25 5—29 5, 29 5—33·5 33·5—37·5,
Racords — 1 1 9

37·5—41·5, 41·5—45·5
11·5 28

45 5—49·5, . .
31·5, 36·5, 30 5, 31·5, 30, 26,
......—77·5—81·5
13·5, 4, 3

Compute the mean, median, standard deviation, quartile deviation.

Ans 55 1, 54 9, 10·33, 7·9

XIV.—

x	f	x	f
3	5	38	79
8	9	43	50
13	28	48	37
18	49	53	21
23	58	58	6
28	82	63	3

XV —Calculate the standard deviation and co efficient of ariation for

Marks	10-20,	20-30,	30-40,	40-50,	50-60,	60-70,	70 80
o of students	5	12	15	20	10	4	2

Represent the data graphically.

(Hyderabad University B A 1946)
Ans. $\sigma=14\ 3$

$cv=85\ 2$

XVI —Compute the Quartile Deviation and co efficient of variation for the data in Q 29 Ex I Also determine the value of quartile deviation graphically

(Indian Audit and Acctts. Exam. 1945).

Ans. 13 54 , 12 24

XVII.—The following is the frequency distribution of percentage butter fat, in samples of milk of individual cows in a herd Calculate any one of the three measures of dispersion state the relative merits of the three as measures of a distribution Percentage butter fat, 2—2'4, 2 4—2 8, 2 8—2 12

8 8—9 2

Frequency 1, 4, 6, 19, 63, 85, 111, 95, 79, 53, 28, 16, 12, 9, 1, 2, 0, 2

Indian Audit and Acctts 1943) *Ans* $\sigma=9\ 5$ *nearly*

XVIII.—Calculate the standard deviation for the following data of 'Difference in age between husband and wife in a particular community.

Difference in years	0—10,	10—15,	15—20	20—25,	25—30,	30—35,	35—40
Frequency	700	507	281	109	52	16	4

(*B Com 1945*) *Ans. 6'77.*

XIX —Compute the standard deviation for Q 27 (Ex I) (*Hyderabad 1945*)

Ans 10 0

XX —Given sa es as Rs 230 393 582 799 1035 fo 5 years in 1927—35

Find the co efficient of variation

(*B Com Supp. 1945*) *Ans 46*

XXI —The following table gives the index numbers wholesale prices of cotton manufactures and wheat in Indi for ten months, from January 1943 to October 1943 Indicat which of the two goods had more variable price

Prices for week ending 19th August 1939 = 100

[Capital June 29, 1944

Index of Cotton 415, 427 437 469, 505 513 493, 426, 4
417

Index of Wheat 252, 332, 312 308 323, 330 346, 371,
380

(*B A Hons 1945*) *Ans Whea*

CHAPTER VI

INDEX NUMBERS

An index number is a statistical device for estimating the lative movements of a variate, in cases where measurements its actual movements are inconvenient or not possible idex Numbers have gained great importance in almost all anches of scientific inquiry We may have index number prices, cost of living, index numbers showing changes in nemployment, production, investment, industrial activity, isiness conditions, health and academic grades etc

The index number will measure fluctuations during tervals of time, group differences of geographical position degree, and it cannot do more than show a general ndency

Construction of Index Numbers—The technique of dex number construction involves the following process

(*a*) Choice of items to be included The items selected ould be representative of the tastes, habits, or require ents of the class of the purchaser concerned The imber of items should be fairly large. To compare anges in the general level of prices in a given period of ne, we should have

(1) Selection of representative Commodities

(2) Selection of representative places for each Commodity

(3) Regular and reliable quotation of prices from the representative places of commodities Wholesale

(*b*) The form of Average to be used

(*c*) The selection of the Base

(*d*) The weighting system. The designation of the degree of relative importance of each constituent item is known as weighting

For simple Index Numbers, the first three methods are sufficient

Choice of Base—With the Fixed Base method, a definite year or average of a period of years is chosen and adhered to for a long time The period selected should be a period of normal conditions and free from fluctuations and disturbances likely to affect the index This Base is taken as 100, the price for this is taken as basic for the purpose of calculating Index Number

Indez Number for a particular year

$$= \frac{\text{Price of the particular year}}{\text{Basic price}} \times 100$$

If the base price is Rs 5, then Index of a particular year when price is Rs 8 is $\frac{8}{5} \times 100 = 160$ This is alsc called a 'Relative' or the Percentage price and the combination of such relatives is called the 'Index Number in the general form Add all the relatives and divide by the number of items to get the general Index or Index N of Prices for the commodities

Example —*Given*

Commodity	*Relatives during the years* 1914 (Base = 100)	1923	1931
Wheat	100	170	72
Rice	100	192	70
Sugar	100	195	95
Ghee	100	187	92
Wood (Fuel)	100	185	92
Gold	100	150	180
Index Number of Prices	$\frac{600}{6}$ = 100	$\frac{1079}{6}$ = 179·8	$\frac{601}{6}$ = 100·1

Here we have used arithmetic mean as the average, it if the geometric mean is to be applied, then tbe Indez umber will be obtained by multiplying the relatives and en taking the 6th root as the number of commodities in is example is 6 (in general *n* th root, where *n* is the umber of commodities), Logarithms should be used for rge numbers Harmonic Mean and Median can also used s an Average but in practice, A Mean and Geometric lean are frequently used, Geometric, being preferred iving better results

Chain Base Method.—With the chain base method, ach year is calculated upon the preceding year as the ase and the results are linked together afterwards as shown n the example

'Statist' Index of Sugar, Tea and Coffee using Chain Base system, Relatives given,

Year	Sugar I	Sugar II	Coffee	Tea	Total	Average	Chain Index
1921	81	77	119	55	332	$\frac{332}{4}=83$	83
1922	62	54	128	82	326	81 5	83 × 10(
	76	70	108	149	403	100 8	10) = 83
1923	104	87	111	100	402	100 5	83 7 × 13
	168	161	87	122	538	134 5	100 = 11

Explanation —Given the Relatives for 1921, 192 1923, for four commodities Find the Average for each ye Take the year 1921 and 1922 Taking 1921, as the ba with chain index 83, construct relatives for 1922, so t for Sugar I we shall have $\frac{62 \times 100}{81} = 76$ nearly. Similarly t other commodities we have 70, 108 and 149 Take t average of these which is 100 8 Multiply this avera by the change index of 1921 and divide by 100 to get t chain index for 1922 which is $\frac{83 \times 100\ 8}{100} = 83\ 7$

Next, find the Relatives for 1923 taking 1922 as t base, we get for Sugar I $\frac{104 \times 100}{62} = 168$ and so on Ta the average of these and multiply it by the chain i of 1922, thus we get chain index for 1923.

$$= \frac{134\ 5 \times 83{\cdot}7}{100} = 112\ 5$$

In this way we proceed further chaining each y with the preceding

Chain base method provides a direct comparison b ween each year and the next, which is more interesting to commercial people than indirect comparison through the medium of a possibly remote base

Weighted Index Numbers

There are two methods of weighting the indices of prices

(*a*) *Weighted Aggregate of actual prices* — When actual prices of the commodities or item are given and also the quantity of each item the quantities produced in some fixed period such as the base year may be used as weights

The index is obtained by comparing the weighted aggregate (total) for the given year to that of the base year
The formula for index number (Base year weighting) is

$\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0}$ where p_1 represents the prices for the current year for which the Index is required, thus for one particular year p_1, for second p_2,

- q_0 represents actual quantity of the base year for each item
- p_0 represents the actual price of the base year for each item

To find the index for a particular year, multiply the price of that year with the corresponding quantity for the base year and add the products for all the items Divide his sum by the sum of the products $p_0 \times q_0$ Multiply the result by 100 to get the Index Number

However, since conditions change, the quantity of the commodities produced in any one fixed period will not a good measure of their relative importance for all periods. To meet this objection a set of weights whic change every year may be used. Thus the quantity p in each given year may be used as weights when constructing the index number for that particular period

The formula can then be written as (current weighting)

$\frac{\Sigma(p_n\, q_n)}{\Sigma(p_0\, q_n)}$ where q_n represents the quantity for particular year and p_n its price. So for first year we can hav q_1 the quantity and p_1 the price, for second year p_2 and q_2 fo quantity and price respectively

If fixed weights are used, the formula will be

$\frac{\Sigma\, w\, p_n}{\Sigma\, w\, p_0}$ where w stands for the weights. If relatives and weights are given the weighted Index No. is obtaine by multiplying the two and dividing the sum of products b sum of weights

The above formulæ are suitable for use with either th fixed or the chain base methods. They are to be multiplied b 100 to get the Index Number

Fisher's Ideal Formula.—It is the geometric mean c the first two formulæ Index for a year

$$= \sqrt{\frac{\Sigma\, q_0\, p_n}{\Sigma\, q_0\, p_0} \times \frac{\Sigma\, q_n\, p_n}{\Sigma\, q_n\, p_0}}\,.$$

This is also called a cross-weight formula.

There are over 150 formulæ for Index Numbers but her we have given the widely used ones, which may be use according to the nature of the data *(b) Weighted Average of Relatives or Ratios, Method*

In this method the price relatives play the part and not the actual prices as in the former method The formula with base year weights,

$$\frac{\Sigma\left[\frac{p}{p_0} \times (p_0\, q_0) \right]}{\Sigma(p_0\, q_0)}$$

Here the price relatives $\left(\frac{p}{p_0} \right)$ are weighted by total expenditure $(p_0\, q_0)$

Through cancellation this formula reduces to

$$\frac{\Sigma(p_n\; q_0)}{\Sigma(p_0\, q_0)}$$

If current year or given year weights are used the formula is

$$\frac{\Sigma\left[\frac{p_n}{p_0} \times (p_n\; q\;\;) \right]}{\Sigma(p_n\; q_n\;)}$$

For one year with p , price and q_1 the quantity

$$= \frac{\Sigma\left[\frac{p_1}{p_0} \times (p_1\, q_1) \right]}{\Sigma(p_1\, q_1)}$$

Index Number Tests—There are two fundamental methods for te ting the consistency of the Index Number

(1) *Time Reversal Test*—Let the index of a year say 1930, computed with base (1928 = 100) be 2 00 reconstructing the Index Number for 1928 with base 1930, the index should be, by Time reversal equal to reciprocal of 2 00, *i e* , $\frac{1}{2}$ = ·5

	1928	1930
Icdex	1 00	2 00
	5	1 00

Cross-multiplying the Index numbers should give a va' of 1'00, since there are reciprocals. The test may be as: If the time subscripts of a price or (quantity) inde number formula, be interchanged, the resulting price ((quantity) formula should be reciprocal of the original formula

Take the formula $\frac{\Sigma\, p_n\, q_0}{\Sigma\, p_0\, q_0}$ and change the time scripts, it becomas $\frac{\Sigma\, p_0\, q_n}{\Sigma\, p_0\, q_0}$. Multiplying the two the resu is not equal to unity (one).

The Arithmetic Average of Relatives is not reversible The result of calculating the current year upon the b year does not agree with the result of calculating the ba year upon the current year. The product of the two i greater than 1 and not equal to 1 as it ought to be.

The simple geometric mean is reversible With the geometric mean, the fixed base and chain base method agree, though it is rather troublesome to calculate th geometric mean

Fisher's ideal Index Number meets the test

Factor Reversal Test.—The index of prices can be obtained by any of the methods, for example, take formula $\frac{\Sigma\, p_n\, q_0}{\Sigma\, (p_0\, q_0)}$.

An index of the quantity of production can be obtained by reversing the position of the price figures (p) with the quantity figure (q) and so it is

$$\frac{\Sigma(q_n\, p_0)}{\Sigma(q_0\, p_0}.$$

The factor reversal test says that

$$\frac{\Sigma\, p_n\, q_0}{\Sigma\, p_0\, q_0} \times \frac{\Sigma(q_n\, p_0)}{\Sigma(q_0\, p_0)} \text{ should be} = \frac{\Sigma\, p_n\, q_n}{\Sigma p_0\, q_0},$$

i.e., if p and q factors be interchanged in a formula the product of the two should be equal to $\frac{\Sigma\, p_n\, q_n}{\Sigma\, p_0\, q_0}$

Fisher's Ideal Index Number

$\sqrt{\frac{\Sigma\, p_n\, q_0}{\Sigma\, p_0\, q_0} \times \frac{\Sigma\, p_n\, q_n}{\Sigma\, p_0\, q_n}}$ transforms itself into

(by interchanging p and q) $\sqrt{\frac{\Sigma\, q_n\, p_0}{\Sigma\, q_0\, p_0} \times \frac{\Sigma\, q_n\, p_n}{\Sigma\, q_0\, p_n}}$

Multiplying the two ideal indices, the result is

$$= \frac{\Sigma\, p_n\, q_n}{\Sigma\, p_0\, q_0}.$$

Fisher's Ideal Index Number is called ideal, as it meets both the tests

Quantity Index Numbers —The Index Numbers can be used to measure changes in quantity groups as well as price changes Index Numbers of this type are applicable to the measurement of changes in business activity, industrial production, *etc* The *method of construction is the same* for Quantity Index Numbers as for Index Numbers of prices. The simplest form is $\frac{\Sigma q_n}{\Sigma q_0} \times 100$

Where Σq_n denotes the sum of the quantities in *any current or given year*

Σq_0 denotes the sum of the quantities in the base year

The weighted aggregate form for measurement of quantity changes is $\frac{\Sigma p_0\ q_n}{\Sigma(q\quad p_0)}$ with base year weights (where p_0 may be the price or some weights), and $\frac{\Sigma p_n q_n}{\Sigma p_n q_o}$ with current or given year weights.

Excercise V

I.—Years		1930,	1931,	1932,	1933,	1934,
Price of wheat per maund	Rs	4	5	6	7	7-8-0
		1940,	1941,	1942,	1943.	
		10	9	10	11	

Find the Index Number (1) by taking 1930 as the Base (2) the average of the first three years as base (3) 1940 as Base

Ans. (1) *100, 125, 150, 175, 187 5, 250, 225, 250 , 275*
(2) *80, 100, 120, 140, 150, 200, 180, 200, 220*
(3) *40, 50, 60, 70, 75, 100, 90, 100, 110.*

II.—Years	1921,		1922,	1923,
Bank Deposit Rs	0000,	34,845	37,194	40,034
	1924,	1925,	1926,	1927
	42,954	46,766	48,882	51,133

Calculate the Index Numbers for the Deposits for each years taking 1921 as base, in round figures.

Ans. 100, 107. 115, 123, 134, 140 and 147.

III.—Find the Index of Bank clearings and of Immigrants from the following data taking the average as the base, in round figures

Year.	*Bank clearings in Million of Rs*	*Immigrants in tens of thousands*
1	49	79
2	40	52
3	25	33
4	35	55
5	35	46
6	34	62
7	28	34
8	34	31

Ans 140, 114, 71, 100, 100, 97, 80, 97,
161, 106, 67. 112 94 126 60 62

✓ IV.—Compare the following prices of Wheat and Coal as to their relative changes for the period 1913—20 Taking 1913 as base find the Index Number for each year for each commodity

	1913,	1914,	1915,	1916,
Price of wheat Rs. per maund annas etc	3 11 6	4-6 6	5-6-0	4-13-0
Coal per ton Rs	6-10 0	6-12 0	6 15 0	7 0 0
	1917,	1918,	1919,	1920
	4 12 6,	5 9 6,	8 3 6,	7 0 0
	7-1 0,	7-3 0,	7-10 0	7 8 0

Ans Wheat 100, 119, 144, 130, 129, 151, 221, 188

Coal 100, 102, 103, 106, 107, 109 115ʻ 113

V.[1]—Calculation of Statist Index of wholesale prices of Minerals

Year average.	Iron shillings and pence per ton		Bars Common £ per ton	Copper Standard £ per ton	Tin straits £ per ton	Lead £ per ton	Coal shillings per ton.	Coal average Export Price shillings per ton
	A	B						
1867—77	69-0	60 0	$8\frac{1}{4}$	75	105	$20\frac{1}{2}$	22	12 5
1913	65 6	58-3	$7\frac{3}{4}$	68	201	$19\frac{1}{8}$	$21\frac{1}{3}$	13 94
1921	168-6	137-4	$19\frac{1}{8}$	$69\frac{1}{8}$	171	$24\frac{1}{8}$	$32\frac{1}{4}$	34'83
1924	96-8	88-2	$12\frac{1}{2}$	$63\frac{8}{18}$	251	$35\frac{18}{16}$	$27\frac{1}{2}$	23'38
1930	76-0	67 0	$9\frac{5}{8}\frac{1}{2}$	$54\frac{3}{4}$	$144\frac{29}{32}$	$19\frac{1}{2}$	$24\frac{15}{32}$	16'64

Taking 1867—77 as the base, calculate the Index Number of Minerals for each year, using arithmetic mean upto two places of decimals.

Ans. 110 65, 180'99, 157'87, 111 95

Hint —First find the relatives.

[1] Connor, Chapter XVI Index Numbers

VI.—In Q V, calculate the Indez Numbers (1) for 1921, taking 1913 as the base and (2) for 1913, with 1921 as the base

(1) 246 4, 246 77 101 95 85 08, 126 54 150 249 86

(2) 40 59, 40 52 98 08 117 54 79 02 66 67, 40 02

VII—In Q V, determine the Index Number for Minerals by taking the Geometric Mean of the Relatives

Ans *106 9, 169 2 151, 111 4.*

VIII—In the solved example on chain base method, find the Chain Index for the years 1924—28 given

1924	93	75	154	96
1925	60	43	165	88
1926	60	44	159	89
1927	62	47	139	84
1928	51	40	146	77

Ans *115 3, 92 8 100, 90 8, 82 6*

IX—In Q VI, find the Index Number of Minerals and test the Index Numbers by the Time Reversal Test

Ans (1, *172 37*, (2) *68 92*,

Product of Indices 1 18 Not Consistent

X—Find the Quantity Index Number for the following data with 1932 as the base

Year	*Quantities* *A*	 *B*
1932	9	7
1933	10 3	9
1934	11 0	6 7
1935	10 5	9 4
1936	12	4 5
1937	9 5	5 4
1938	8 9	9

Use the first formula given in Quantity Index Nos

Ans (in round figures,
100 121, 111, 124, 103, 93 and 112).

XI —Explain the methods used in constructing the Index Numbers of wholesale prices, or of the cost of living giving illustrations Define an Index Number and explain the role of 'weights' in the construction of an index of the general price level

(B A Hons , M A 1942, 1943, 1945 , B. Com 1945)

XII —Explain with illustrations what is understood by an Index Number ?

Discuss the relative advantages of (1) Arithmetic Mean, (2) Geometric Mean, (3) Harmonic Mean, in the construction of an Index Number

(Indian Audit and Accounts, 1941)

XIII —Find the cost of living Index Number for the working classes from the data in Q XIII and Q XIV

Articles	*Quantity Consumed in 1914 (Base) q_0 in Crores*	*$p_0 \times q_0$ for 1914, Rupees in Crores*	*$q_1 p_0$ p_1 is price fo 1924*
Pulses	13 maunds	60	70
Cereals	108 ,,	583	746
Food Articles	46 ,,	381	728
Firewood and Coal.	50 ,,	60	101
Clothing	88 Pounds	53	121
House Rent	Rs. 10 per month	113	187

XIV —

ommodities (1)	*Annual Expenditure in 1914* (2) Rs	*Weights Assigned* (3)	*Relatives for 1931* (4)
ice	5	10	70
ajra	5	10	65
/heat	40	80	72
ulse gram	10	20	60
,, Arbar	15	30	80
/ood	5	10	92
ugar	2/8	5	95
alt	1	2	90

(1)	(2)	(3)	(4)
Ghee	10	20	92
Oil	5	10	87
House Rent	6	12	120
Potato	2/8	5	95
Gold	Nil	0	180
Cotton Cloth	15	30	96
Cloth	5	10	95
Brass	2/8	5	90
Oil	5	10	110

Ans 83 nearly

XV —

	Prices			*Quantities*		
Crops	*Basic Year Price po*	*(1927)*	*(1928)*	*Basic Year qo*	*1927*	*1928*
1	64 2	72 3	75 2	26 2	2763	2819
2	119 8	111 5	97	831	878	915
3	39 8	45	40 9	1247	1182	1439
4	57 5	67 8	55 2	185	266	357
5	141 4	96 5	53 6	354	403	465
6	10 9	19 6	18	8989	6478	7239
7	1410	1135	1227	86	106	93
8	18 2	21 2	20	1298	1211	1374

Find the Index Numbers by

	1927	*1928*
(1) Base year weighting	110 5	105 7
(2) Current year weighting	105 7	101 1
(3) Fisher s formula	108 1	103

XVI —Use formulae (*b*) to find the Index for 1927 in Q XV

Ans 110 5, 114 5

XVII —Apply Tests to XV

XVIII —It is desired to find the difference in the cost of living in the years 1939 and 1943 in the case of (i) Clerks (ii) industrial labourers in a big industrial town

Explain fully the necessary procedure to be adopted

(B Com 1945 Supp)

XIX —Distinguish between Fixed Base and Chain Base methods of constructing Index Numbers giving examples

Describe the various methods of weighting the index numbers of prices

How can the Index of Indian industrial activity be constructed ?

(Indian Audit & Accounts Exam 1945)

XX —What is an index number ? What are (1) time reversal test and factor reversal test ? State their use

(C St & M A 1945.)

CHAPTFR VII

ANALYSIS OF TIME SERIES

The analysis of Time Series involves the description and measurements of the various movements or changes as they come in the series during a period of time The characteristics of a time series are to be found in its trends and fluctuations which are described here very briefly

1 Secular Trend or the long time growh or decline, existing within the data It is a smooth, regular and long term movement of a statistical series Most series of economic statistics exhibit definite trends Such a trend may be constant in direction or may change direction at a constant rate. Thus the volume of production or sales of business house over a period of years shows a fairly regular growth. The same is the case with population of a country

2. Fluctuations in time series may be regular or irregular *Regular fluctuations are (1) long term fluctuations (i e. the* Trend) (2) Periodic or moderately long period fluctuations, (3) Short term fluctuations or seasonal variations, which are more or less regular movements within the twelve month period and due to the changing seasons, consumption and production of commodities, interest rates, etc are marked by seasonal swings repeated with minor variations year after year

3 Cyclical movements or the swing from prosperity through recession, adversity, recovery and then on to prosperity again One cycle is said to be completed when beginning with a peak, the falling curve reaches a minimum point and then rising again reaches the next peak This is the case with price fluctuations

4 *Residual, accidental* or random Variations, including unusual disturbances catastrophic or unexpected events such as wars, disasters, famines, strikes, floods

Measurement of a Trend

The following methods are commonly used *to measure* trends :—

(1) Freehand drawing (2) Semi average (3) average (4) Fitting a curve by least squares, which explained in the next chapter on 'Curve Fitting (VIII)

1 *Freehand drawing*—First of all draw the graph o the given time series, with the time along the horizontal axis Draw a smooth freehand line (or curve approximately carefully in such a way as to describe what appears to be long period movement

2 *Semi average method*—In this method break the d into two equal parts and mark the middle years of each (if th number is odd, taken two parts approximately equal) Tak the average of each part Plot these averages at the mid points of their respetive periods Join the two points drown, this line will show the trend

3 *Moving Average Method* is used for smoothin fluctuations in curves and to exhibit a trend with the hel of averages in years Smoothing brings out tendencies T moving Average may be for three five six seven years and so on according to the size of the data For thre years moving average take the average of the first three year and place it against the middle year of the three Leav the first year and then take the average of the next thre years and place it against the middle of these three year Proceed in this way taking the average after leaving on preceding year Then plot these moving averages alon with time series graph This will be a moving Averag graph showing the Trend For a five year moving averag take the a [illegible] the f[illegible] five [illegible]

ıddle year Then take against the next five years leaving the rst year and place it in the middle year of these Pro ed in this way and draw the curve For a moving verage of even years say four take the average of the rst four years and place it against the middle *i e*, between cond and third year Leaving the first year, take the verage of the next four years and place in the middle these Proceed in this way and then draw the Trend aph A seven year cycle may be eliminated by means a moving Average based upon a period of 7, 14 years he greater the number of years the smoother the curve

Example 1 —Years	*Values*	*3 Year Moving Total*	*3 Year Moving Average*
1921	8		
1922	6	21	7
1923	7	24	8
1924	11	30	10
1925	12	37	12 3
1926	14	41	13 66
1927	15	48	16
1928	19		

When an even number of items is included in the oving average, say six the centre point of the group es between two years It is necessary to adjust these x year moving averages so that they coincide with ars Take a two years moving average of the six ars average The resulting average is located between e two six year moving average values and, therefore, incides with the years The final result is said to be e six year moving average centred

Example 2.—Year. (1)	*Values.* (2)	*Six year Moving average.* (3)	*Two year Moving Total of Col 3.*	*Six Mov Av.* *Cє*
1924	16			
1925	17			
1926	25			
		32		
1927	35		68 66	34'33
		36'66		
1928	46		78 82	39'41
		42'16		
1929	53			
1930	44			
1931	50			

The Moving average is quite simple for calclatı and especially useful in making approximations of general movements in a series particularly eliminaı a large part of a cycle that is rather regular. 1 average cannot be brought up-to-date, as, depend upon the number of items included, the last point in trend occurs a few years before the end of the data.

Moving Average and seasonal variations —Mo. Averages provide a useful method for isolating sea Variations First of all take the moving average for months, centred (adjusted by two month-moving for all the years. Express the original data as ages of the corresponding moving averages. Take average (arithmetic or median) of the percentages for ι month (dividing by the number of years for Mean).

These will be the Indices of seasonal Variation ach month The average of the 12 Means for 12 months ught to be 100 otherwise the Means may be adjusted o as to have the average 100 (e g See Exercise VI, 8)

There is a simpler method for measuring the seasonal ariations by taking the averages, which can be used when ne general trend is fairly steady or has only a slight upward r downward slope otherwise adjustment has to be made or the Trend The simple avera e method may be described s follows An average value is obtained for each month nd then a final average of all the monthly averages dividing y 12 By subtracting this mean of means, from the verage figures for each month the seasonal Variations for ach month are obtained (See Exercise VI, 7) At least ree or preferably more years figures should be taken

Besides the methods explained above, the methods of ık Relatives and 'Ratio to trend' are used which are ther complicated The ratio to trend method measures e seasonal Variation and in addition the combined cyclical d residual Variations and depends on fitting a trend line the data

The Link Relative method, is based on 'link Relatives' r which we express the value for each month as a rcentage of the previous month. The resulting percentages e called link relatives Median is used as the average

Exercise VI

1 Draw a freeband Trend for the following time series

1910	1911	1912	1913	1914	1915
810	890	780	784	846	775
1916	1917	1918	1919	1920	1921
816	820	875	750	807	750
1922	1923	1924	1925	1926	1927
36	807	735	783	780	760
1928					
720					

2 Draw the graph for the data in Q 1 and also the graph of t ree years moving Average

3 Draw the Trend by the semi average method from the following data

	1914	1915	1916	1917	1918	1919
Values	16 0	18	25 3	35 3	46 6	35 2

1920	1921	1922	1923	1924	1925
44 6	50 9	53·6	64 5	70	79
1926	1927	1928	1929	1930	1931,
89 5	97 2	105 92	119	119 62	114 5

Hint —Take up to 1922 first half with 1918 as middle year and find the Average S milarly for the other half with 1927 us middle year

4 Given the Index Number of food prices in the Punjab (1873—82—100)

Years 1861 1862 1863 1864 1865 1866 1867 1868
109 67 59 71 83 83 94 126
1869 1870 1871 1872 1873 1874 1875 1876
169 119 93 100 82 84 77 72
1877 1878 1879 1880 1881
78 134 151 125 111

Fınd five yearly average and plot

Ans 78 73 78 91 111 118 120 121 113 96 87 83 79 89 102 112 and 120

5 Fınd the n ne vears Movıng Average for the series

9 7 5 2 4 9 10 9 8 6 4 7 11 13 11 9 8 5 10 13 15 12 10 8 6 11 12 16

Ans 7 6 7 6 3 6 6 7 6 8 6 8 8 8 7 8 6 8 2 8 7 9 7 10 6 10 7 10 3 10 9'7 10 10 8 and 11 4

6 Fınd the s x year Movıng Average for Q 3 and draw the Trend ınd cated

7 Fınd the seasonal Variatıons us ng the Sımple Average Method from the followıng data

	Jan	Feb	Mar	Apr	May	June
1930	50	42	38	41	36	42
1931	45	43	45	47	44	40
1933	41	40	34	37	39	41

	July	Aug	Sept	Oct	Nov	Dec
1930	40	42	41	48	50	50
1931	52	50	48	47	46	43
1933	41	41	39	39	48	46

Sol —Total of monthly averages ıs 518 7

Mean of means = 43 2 and the seasonal Varıatıons for each month are 2 1 −1 5, −4 2 −1 5 −3 5 −2·2 1 1 1 1 − 5 1 5 4 8 3 1

8 Determine the seasonal Variations using the average method from the following data (Mills)

Months	1925	26	27	28	29	30	31	32	33	193
January	655	728	820	706	696	848	859	891	920	94
February	753	687	776	685	757	854	908	906	932	95
March	842	696	848	691	818	916	916	926	960	99
April	873	721	730	706	716	941	874	932	966	96
May	897	759	862	760	776	975	895	971	1018	100
June	918	796	896	762	831	1012	905	992	1052	10
July	970	887	901	750	813	985	881	975	1037	97
August	962	892	969	810	853	1042	969	1073	1106	107
September	956	960	967	842	925	1037	1037	1074	1140	10
October —	925	967	1005	932	978	1070	1091	1107	1184	110
November	819	807	884	764	957	964	976	1024	1042	92
December	719	758	755	681	832	827	869	925	858	8

Sol —*First find the 12 monthly moving average centred* these will be from July 1925 (860 5) upto June 1934 (991 for this June) The seasonal variations will be 91 6 92 1 *95 8 92 8 98 6 101 6 102 4 107 9 111 1 115 101 7 89 4*

9 Explain what is meant by (*a*) the secular trend and (*b*) seasonal fluctuations in a time series Indicate briefly the procedure of estimating these

(*Indian Audit and Accts Service 1941*)

10 Describe the various types of fluctuations in a Time Series and explain the procedure of isolating them or Write an essay on Time Series

(*M A 1941 1942 and 1943*)

11 What is a trend' and how is it measured? Use e method of Moving Averages to dete-mine the trend in e following Series showing index Numbers for values of nports into India during 1914—1928

87, 62, 47, 42, 45, 57 96, 97 84 79, 77, 80, 92, 106 and

(*M A 1942*)

12 Show how Trends are measured

(*B Com 1945*)

CHAPTER VIII

METHOD OF LEAST SQUARES, CURVE FITTING AND TRENDS

Curve fitting is an important subject from both theoreical and practical point of view It is the representation of relationship between two variables by simple algebraic expressions The chief method for fitting of curves to a given data is by means of the least *square method* According to this method, we suppose the curve best fitted to be of the form

$$y=a+bx+cx^2+dx^3+ex^4+ \cdot \dots$$

If a straight line is to be fitted the equation takes the form $y=a+bx$ (two unknowns a and b)

For a second degree curve or second order parabola the equation takes the form $y=a+bx+cx^2$ (three unknown a, b and c)

For a third degree curve (a third order parabola) the equation takes the form $y=a+bx+cx^2+dx^3$ (four unknc a, b, c and d) and so on

General procedure —

(*a*) Write down the type of the equation to be and substitute the values of x and the corresponding y the equation

(*b*) Form Normal equations for each unknown. T Normal equation for the unknown 'a' is obtained by mul plying the equations by the co efficient of 'a' and addi The sum will be

(1) $\Sigma y=na+b\Sigma x$ for a st line.

(2) $\Sigma y=na+b\Sigma x+c\Sigma x^2$ for a second degree curve

(3) $\Sigma y=na+b\Sigma x+c\Sigma x^2+d\Sigma x^3$ for a third degree

where n is the number of items.

(*c*) Form Normal equation for the unknown b multiplying the equations by the co-efficient of b (which is and add. The sum will be

(1′) $\Sigma xy=a\Sigma x+b\Sigma x^2$ for a st. line.

(2′) $\Sigma xy=a\Sigma x+b\Sigma x^2+c\Sigma x^3$ for a second degree curve

(3′) $\Sigma xy=a\Sigma x+b\Sigma x^2+c\Sigma x^3+d\Sigma x^4$ for a third curve

(*d*) Form Normal equations for the unknown c, multiplying the equation by the co-efficient of c (which is and add. The sum will be

(2″) $\Sigma yx^2=a\Sigma x^2+b\Sigma x^3+c\Sigma x^4$ for second degree curve.

(3″) $\Sigma yx^3=a\Sigma x^3+b\Sigma x^3+c\Sigma x^4+d\Sigma x^5$ for third curve (*e*) Form Normal equation for d by multiplying the equations by the co efficient of d (i.e., x^3) and add. We get

(3‴) $\Sigma yx^3=a\Sigma x^3+b\Sigma x^4+c\Sigma x^5+d\Sigma x^6$.

In general, the set of Normal Equations for the curve

$$y=a_0+a_1x+a_2x^2+\dots+a_kx^k\text{, are}$$

$$\Sigma y=a_0n+a_1\Sigma x+\dots+a_k\Sigma x^k$$

$$\Sigma xy=a_0\Sigma x+a_1\Sigma x^2+\dots+a_k\Sigma x^{k+1}$$

$$\Sigma yx^2=a_0\Sigma x^2+a_1\Sigma x^3+\dots+a_k\Sigma x^{k+2}$$

$$\Sigma yx^k=a_0\Sigma x^k+a_1\Sigma x^{k+1}+\dots+a_k\Sigma x^{2k}$$

The number of Normal Equations will be the same as the number of the unknowns Solving these equations simultaneously we get the values of the unknowns These will be the most plausible or most possible values for these unknown quantities satisfying the set of equations obtained by substituting the various values of x and y Such equations are known as the equations of observation Putting the values of the unknowns in the equation, to be fitted, we get the required equation which represents the curve fitted to the data

When the number of Normal equation is more other methods such as of (1) Determinants, (2) M equal co efficients, (3) Gauss's method, (4) D method may be used The Normal equations for a line from the general procedure are (1) and (1′) for a bola (2) 2′, 2″, for a third degree curve, (3) 3′, 3″, 3 which can be easily solved simultaneously When we solved the equations for a, b, c and d, put the values in t respective equations to get the best fitted curve

To fit a st line to the given values of x and y

	x	y	xy	
	1	3	3	
	2	4	8	
	3	6	18	
	4	5	20	
	5	10	50	
	6	9	54	
	7	10	70	
	8	12	96	
	9	11	99	8
Sum of Σ	45	70	418	28

The equations of observations are obtained by putti the value of x and y in $y=a+bx$, they will be,

$3=a+b$, $4=a+2b$ and so on Items here are 9, so

The Normal Equations are $\Sigma y=na+b\Sigma x$,

$$70=9a+45b$$

$$\Sigma xy=a\Sigma x+b\Sigma x^2=418=45a+285b$$

Solving these two equations, we get the *most plau* values of a and b as $a=2{\cdot}11$, $b=1{\cdot}13$ The equation to t wn on the graph paper In this way the parabola of ond order and third order or any other curve can be best ed after forming the Normal equations and solving m

Trend and the curve fitting

The Method of least square is applicable for the deterination of the Trend In a Time Series, where the period given and equi spaced values corresponding to the periods represented by y, the Time, years etc, are assigned numbers 0, 1, 2, 3, 4, 5, 6, 7, and they are taken x The starting year to which the number 0 is assigned, is own as the Origin Year Here n will denote the total mber of years The rest of the process is the same as plained above Form Normal equations and solve them the usual way If the best fit is a line, this will be a linear rend otherwise non-linear Trend The line is also called e Least Square line The straight line trend does not tisfactorily describe the trend of data which have a vary g rate of growth In such cases a parabola may be ited

The trend values (y) for the various years may be tained by substituting the appropriate values of x from e numbers 0, 1, 2, assigned to each year, in the equan obtained for the trend These can be plotted on the raph paper to draw the curve.

Short method for trends—If the number of years is dd, take the middle year as origin year and assign 0 to it years and 1, 2, 3   to the succeeding years, so that will be zero Thus if the years are 1919, 1920, 1921, 19 1923, the middle year is $1921=0$, the preceding years 19 1919 will be -1, -2 and the succeeding years 1922, 19 will be 1, 2 so that $\Sigma x=0$ In this way, the working simplified, and the simplified Normal equations will $\Sigma y=na$, and $\Sigma xy=b\Sigma x^2$ for a linear trend Similarly f the parabolas Of course the origin year will be the mid year and not the year of start as in the general case.

For even number of years, the middle term some difficulty To make $\Sigma x=0$ say for a series of sixt years 1926, 1927 - 1933, 1934, 1935——-1941, take two middle terms (1933 and 1934) as $-·5$ and $+·5$ a the other years as with a difference of 1 $-7·5$, -6 $-5·5$, $-4·5$, $-3·5$, $-2·5$ $-1·5$, $-·5$, $·5$, $1·5$, $2·5$, $3·5$, 4 $6·5$, $7·5$, so that the sum is zero the origin being middle of two centre years If decimals are to be avo trend equation may be obtained by working in terms half years, doubling the above assigned values and taki them for x numbers The rest of the process is the sar as explained above

The constant 'a' in the trend equation defines t trend value in the year taken as origin If the ann data employed in the fitting process are averages of twe monthly values, 'a', measures the trend value for a mon centring at the middle of the year covered by the annu values

Graphs of time series on logarithmic scale have be

nding to a time series are used substitute (logy) in
ace of y in the equations and proceed in the same way
obtain the logarithmic trend

rves of the type $y=ax^b$ and $y=ab^x$

Occasionally neither the straight line nor the para-
la will describe the trend of a particular series The
rves of the type $y=ax^b$ and $y=ab^x$ may describe the
nds The equation $y=ax^b$ reduces to log $y=\log a +$
log x The Normal equations are formed by changing
into log y and x into log x The remaining process is the
ne as in the case of a straight line trend The
uations are to be solved for log a and b Similarly
e exponential curve $y=ab^x$ reduces to log $y=\log a+x$
; b and can be likewise treated There are some expon-
tial curves of importance for trend purposes One of
e more important curve is known as Gompertz curve,
iose equation is $y=ab^{c^x}$ Its use in the analysis of
onomic statistics has been based upon the ground that
ere is a general law of growth characteristic of popu
tion increase and that this kind of growth is found in
dustries whose products are a direct function of the growth
population

A somewhat similar curve of growth is the logistic
rve employed in forecasting population growth A form
this curve adapted as a measure of trend is given by

$$1$$

Exercise VII

I —Fit a straight line to the following data ·

x 20 5 10 15 10

y 24 15 17 22 12

Ans $y = 73x + 9\ 23$

What is the line if one more item is added

x 24 *Ans.*

y 30

Ans $y = 86x + 7{\cdot}96$

II.—x 7, 6, 7, 8, 8, 8, 9, 9, 10

y 5, 5, 4, 3, 4, 5, 4, 3, 3

Ans $y = -5x + 8.$

III.—Fit a st line and parabolas of the second and third orders to the following

x 0, 1 , 2 , 3 , 4]
y 1, 1·8, 1·3, 2 5, 6 3]

Ans. $2{\cdot}58 + 1{\cdot}13x$

$$y = 1\ 4 + 1{\cdot}13x + 5x^2$$

$$y = 1\ 4 + 025x + {\cdot}5x^2 + 32x^3$$

IV.—

x	y	x	y
63	40	3193	290
223	1565	2238	259
755	188	1228	231
165	78	2695	255
1535	315		

Fit a quadratic parabola (*i.e.* of second order).

Ans. $y = 48{\cdot}33 + {\cdot}238x - {\cdot}00005x^2.$

V —Fit a parabola of second degree to the following data and draw it

Years	y	*Years*	y
1910	81	1930	134
1915	84	1935	148
1920	88	1940	170
1925	104		

Upon the hypothesis that y continues to increase during the next decade according to this trend, extrapolate or estimate the number for 1950

> *Hint* —The years are equispaced with a difference of 5, so x may also be taken as 0, 5 10, 30

with a difference of 5 The middle year may be taken as origin

Ans $y=78\ 6+ 68x+ 082x^2$

For 1950 (when $x=40$) $y=237$ nearly.

VI —Find the most probable values of x and y from $x+y=3\ 01$, $2x-y=\ 03$ $x+3y=7\ 02$, $3x+y=4\ 97$

(M A , 1942)

Ans *99 and 2 08*

VII —Fit a second degree parabola to the data and plot it

x	1	2	3	4	5	6	7	8	9
y	2	6	7	8	10	11	11	10	9

Ans $y=\ -929+3\ 52x-\ 267x^2$.

VIII —Given the data

Year	1932	1933	1934	1935	1936	1937	1938	1939
x	1	2	3	4	5	6	7	8
[illegible]	722 4	713 0	763 7	770 0	901 1	901 5		[illegible] 3

Fit a curve of the tyoe, $y = ab^{x}$

Ans. $\log y = 2\ 75645 + {\cdot}03044x$

(With origin 1931).

Or $y = 570\ 8(1{\cdot}0726)^{x}$.

IX.—In the following data, S denotes son's stature, and F, father's, in inches

S	65·7	66 8	67·2	69·3	69 8	70 5	70 9
F	62	64	65	69	70	71	72

Fit the relation $S = a + bF$, by determining most probable values of a and b *(Aligarh, 1943 M A)*

Ans. $S = 33{\cdot}351 + {\cdot}522F$.

X.—Given the data for (1920—1938)

805, 895, 785, 784, 846, 775, 816, 823, 874, 750, 807, 750, 736, 807, 734, 785, 784, 765, 715.

Fit a st line with (1919 as origin)

Ans $y = 839\ 4 - 4\ 8x$

Hint.—Take values for years 1920, 1911 1938 as 1, 2, 3, 4, 5, ...

XI.—Given Annual Production of wheat (1926—1940) in millions of maunds in a country, find the linear trend with (1) 1926 as origin, (2) 1933 as origin

y = 111, 143, 143, 134, 138, 55, 74, 129, 150, 140, 145, 160 210 225 229

Ans (1) $y = 95\ 15 + 7{\cdot}2x$. (2) $y = 145{\cdot}753 + 7{\cdot}23x$

Is the production trend figure the same for 1933 by the two methods ?

[illegible]

XII —Fit a st line to the data for 1926—1941

900 1022, 1040, 1080, 1111, 1137, 1176, 1260, 1363, 1420, 1484, 1590 1727, 1828, 1890, 1895

$y = 1370\ 12 + 34\ 2x'$ where x' is in $\frac{1}{2}$ year (double of $-7\ 5, -6\ 5, -5\ 5 \quad 7\ 5)$

XIII —For data of Exports in crores for a country, (1925—1940), fit a second degree parabola with 1925 as origin

2 6, 8 5, 10, 13 3, 9, 15 3, 12 7, 13 8, 20, 28 3, 30 6, 42 5, 44 3, 53, 62 and 65 6.

Find the trend values for 1935 and deviation of the actual from the trend

Ans $y = 7\ 21 - 51x + 304x^2$ *: 32 53 nearly.*

Hint —Actual value for 1935 = 30 6 and deviation from trend = 30 6 − 32 53 = −1 93

XIV —Data of Index Numbers (1915—1927)

114, 110 100 110, 100, 125, 115, 125, 135, 120, 115, 125, 110

Determine the ordinates for the trend of the cubic parabola (1914 as origin)

Ans 112, 109, 107, 106, 109, 114, 118, 123, 126, 127, 125, 120 and 110 (nearly).

XV —Fit a curve to the population of India of the form $y = ab^x$ given in crores,

1871	26·16	1901	29 64
1881	26 68	1911	31·53
1891	29·23	1921	31 1

Hint —Take x as 1 2 . 6 x

XVI.—Find the most plausible values of x, y and z from $x-y+2z=3$, $3x+2y-5z=5$,

$4x+y+4z=21$, $-x+3y+3z=14$.

(*M.A. 1943*). *Ans.* *2 5, 3 5, 1 9*

XVII —Form normal equations and solve

$x+2y+z=1$, $2x+y+z=4$, $-x+y+2z=3$, $4x+2y-5z=-7$.

(*M A 1945*) *Ans* *1 16, − 74, 2 08.*

XVIII —Explain what is meant by (*a*) Secular Trend, (*b*) Seasonal Variations Show how the method of curve fitting is used in the measurement of a Trend

(*Indian Audit & Acctts. Exam 1945*)

XIX —A manurial experiment on paddy gave the following results

Dose of measure in lbs (n) 0 200 400 600

Yield per acre in lbs (y) 1544 1898 2133, 2327

Plot the relationship between the two and use the le square method to fit a parabola of the second degree to represent it

Ans. $y=1547{\cdot}9+378\ 4x-40x^2$ (*C. St & M A 1945*)

CHAPTER IX

CORRELATION AND REGRESSION

So far we have been dealing with the problems which arise from variation in a single variable We will now deal with the simultaneous variation of two or more variables. Methods of measuring the degree of relationship existing between two variables have been chiefly developed by Galton and Karl Pearson It is often desirable to observe and measure the relationship (association), between two or more statistical series For instance it may be desirable to know whether there is relaticnship between changes in the cost of living and changes in wages, the amount of electrical current passed through a solution and the amout ef substance deposited by electro chemical reaction, prices of food grains and rainfall,

When two quantities are so related that the fluctuations in the one are in sympathy with fluctuations in the other, so that an increase or decrease of the one is found in connection with increase or decrease of the other, (or inversely), the two quantities are said to be correlated and the correlation is said to be simple in case of two variables.

Correlation may be direct or positive, if an increase, or decrease in the values of one set is associated with increase or decrease of the other set. If the increase or decrease is associated with decrease or increase of the other, correlation is said to be inverse or negative

Let there be two series x and y to be represented graphically.

Take the items in x series along the axis of x, and the corresponding items in y series along the y-axis The diagram so formed will be a dotted one and scattered, showing some relationship Such a diagram is called a Scattered Diagram

Co-efficient of Correlation —The numerical measure of correlation is called the co efficient of correlation, denoted by r, which lies between 1 and -1 If $r=1$, correlat is said to be perfect If $r=o$, there is no correlation at all Correlation is said to be Null

The following formulæ are used to find the co of correlation

1 Ungrouped data, for x and y series

$$r=\frac{\Sigma\ d_x\ d_y}{n\ \sigma_x\quad \sigma_y}$$

where d_x stands for deviations of the items in x from the arithmetic mean of the x series

d_y stands for deviations of the items in y series, the arithmetic mean of y series

σ_x the standard deviation for x series,

σ_y the standard deviation for y series

n the number of items

The formula is known 'Product Moment' formula to Pearson

It gives a measure of the intensity of the association of the pairs of observations

Suppose it is required to find relationship between the x and y series given by :—

x	y	d_x	d_y	$d_x\ d_y$	d_x^2	d_y^2
28	27	-2	2	-4	4	4
27	20	-3	-5	15	9	25
28	22	-2	-3	6	4	9
23	18	-7	-7	49	49	49
29	21	-1	-4	4	1	16
30	29	0	~~0~~ 4	0	0	16
31	27	1	2	2	1	4
36	29	6	4	24	36	16
35	28	5	3	15	25	9
33	29	3	4	12	9	16
390	250			123	138	164
means · 30 —	25					

Means of the two series are 30 and 25

$$\Sigma\, d_x\, d_y = 123$$

$$\sigma_x = \sqrt{\frac{\Sigma\ (d_x)^2}{n}} = \sqrt{\frac{138}{10}} = 3\ 715$$

$$\sigma_y = \sqrt{\frac{\Sigma\ d_y^2}{n}} = \sqrt{\frac{164}{10}} = 4{\cdot}05$$

$\therefore$ Co efficient of correlation $r = \dfrac{123}{10 \times 3{\cdot}715 \times 4\ 05} = {\cdot}817.$

Since r must be between 1 and -1, it is evident that we have a fairly high degree of correlation

2. Instead of finding the arithmetic mean we can short cut method by taking the Provisional mean and the formula

$$, \quad r = \frac{\frac{\Sigma D_x D_y}{n} - \frac{\Sigma D_x}{n} . \frac{\Sigma D_y}{n} .}{\sigma_x \sigma_y}$$

where n is the number of items and D_x, D_y are the deviations of the respective items from that Provisional Mean.

3 *Correlation for grouped data* —When the x seri and y series are given as frequency distributions, they c- be placed in the form of a Table with one series on horizont- side and the other vertically as shown in the follow example. The table is called 'Correlation Table'. T formula for a correlation table is

[illegible]

$$r = \frac{\frac{\Sigma f D_x D_y}{n} \quad - \frac{\Sigma f D_x}{n} . \frac{\Sigma f. D_y}{n}}{\sigma_x \sigma_y}$$

where D_x denot esthe deviations of the Central Value s f.

the Provisional Mean in x series.

D_y denotes the deviations of the Central Values f

the Provisional value in y series.

f denotes the corresponding frequencies.

n denotes the total number of frequencies.

The whole working will be clear from the followin example -

Example 2 —Correlation Table showing age in years of the students and the Marks obtained

	D_x	17—21 = −4	−2 − 1	0 0	2 +1			
y Series Marks.	*Age in years* *x* ⟶ *Central Values*	16—18 17	18—20 19	20—22 21	22—24 23	*Total of Frequencies for y Series*	D_y	
10—20	15	2	1	1		4	15 − 35 = −20	−2
20—30	25	3	2	3	2	10	−10	−1
30—40	35	3	4	5	6	18	0	−0
40—50	45	2	2	3	4	11	10	+1
50—60	55		1	2	2	5	20	+2
60—70	65		1	2	1	4	30	+3
Total of frequencies for *x* Series		10	11	16	15	52		

(*M.A.* 1939 *and* 1943, *M.A. Aligarh* 1941)

We are given two frequency distributions denoted by x series and y series, x series being horizontal Column D_y gives the deviations from the Provisional Mean 35, (corresponding to Maximum Frequency 18) of the Central Values of y series

D_x gives the Deviation of lhe Central Values of x series from the Provisional Mean 21 (corresponding to the maximnm frequency 16)

$$\Sigma f D_x D_y = 2\times -4\times -20+3\times -4\times -10+3\times -4\times 0 +2\times -4\times 10+1\times 40+2\times 20+0-2\times 20+1\times -40+1\times -60 +0+0-\underline{2\times 20}+0+4\times 20+2\times 40+1\times 60=320$$

$$\Sigma f D_x = 10\times -4+11\times -2+0+15\times 2 = -32$$

$$\Sigma f D_y = 150$$

32
32
964

$$\sigma_x = \sqrt{\frac{\Sigma f (D_x)^2}{n} - \left(\frac{\Sigma f D_x}{n}\right)^2} = \frac{\sqrt{12704}}{52}$$

$$\sigma_y = \frac{\sqrt{61100}}{52}$$

∴ Co-efficient of Correlation

$$r = \frac{\frac{320}{52} - \left(-\frac{32}{52}\right)\left(\frac{150}{52}\right)}{\frac{1}{52}\times\sqrt{12704}\times\frac{1}{52}\times\sqrt{46110}} = 28$$

If the arithmetic mean is to be used for calculatior then

$$r = \frac{\Sigma d_x d_y}{n\sigma_x \sigma_y}. ✓$$

Correlation of Time Series

Correlation for long term fluctuations—When it is desired to measure the correlation in long term fluctutions, for economic and commercial data, Pearson's formulaa is used *viz.* $r = \frac{\Sigma d_x d_y}{n\sigma_x \sigma_y}$ one series say x series is called 'Subject' and the other series or y Series the Relative. The subject is applied to the more important series.

Correlation for short term fluctuations

To study relationship existing in short term fluctuations instead of using the deviations of the items of the Relative and the Subject, from the arithmetic average, we calculate the deviations from the Trend.

The moving average of the Index Numbers of the two factors is calculated and the deviations of such figures from the moving average of the Indices will be the measure of standard deviation in each of the cases.

The rest of the method of calculation is the same as shown above.

Co efficient of Concurrent or Concomitant Deviations

Various difficulties arise when using Pearson's formula in connection with time series subject to short term fluctuations and a co efficient of correlation called 'co efficient of concurrent deviation' exists which gives a simple and easily calculated co-efficient

If it is required to know only whether two series move in the same direction or if one series moves in the opposite direction from the other, the concurrent or concomitant deviations may be used as a basis for measurement. Concurrent deviations are those deviations that are in the same direction for corresponding items in each series.

	Subject		*Relative*	
	Out put of coal Tons.	*Deviations from preceding months or ' First Differ ences.*	*Unemp loyed in coal Industry 000's*	*First Differ ences*
January	18·5		260	
February	19·2	+·7	265	+5
March	19·3	+·1	261	−4
April	18·5	−	274	+
May	17·2	−	292	+
June	15·9	−	357	+
July	15·1	−	330	−
August	16·6	+	306	−
September	17·9	+	258	−
October	17·6	−	280	+
November	18·2	+	250	−
December	19·3	+	225	−

The formula for co efficient of concurrent deviations

$$= \pm\sqrt{\pm\frac{2c-n}{n}}$$

where n is the number of items, c = number of concurrent deviations

If the expression $2c-n$ is negative, we must insert a minus before and after $\sqrt{}$ Hence the general expression $\pm\sqrt{\pm}$ is done to avoid an expression containing the root of a negative quantity.

In the above table, $n=11$, and $c=2$ (as there are two concurrent cases, in February and July)

Co efficient of concurrent deviation

$$= \pm \sqrt{\pm \frac{2 \times 2 - 11}{11}}$$

$$= - \sqrt{-(-\tfrac{7}{11})} = - \sqrt{636} = - 8$$

This co efficient is influenced only by the direction of the deviation and not by the magnitude

This is of no use for long term fluctuations Its principal value is that it indicates the driction of the movements of one series in relation to the other

Ratio of Variation and line of Regression —There may exist almost perfect correlation when two series move, but the proportional movements may be very different. In many cases a measure of this proportional variation can be usefully employed, and the proportional variation for both series having been obtained, comparison of the two by means of a ratio gives us the ratio of variation

When the movements are regular, the Ratio of variation is obtained as follows —

Take the deviation of the relative items from the mean at each date and divide it by the corresponding deviation of the subject Add the quotients so obtained and divide by the number of quotients

Since economic and social series are irregular, it has been found it practice that the Ratio of variation is best determined by graphical method as follows —

Plot the Index Numbers (or first convert if index num- are not given) with subject on the vertical and the of points widely scattered. A line is drawn through the scattered points most nearly approximating the general trend of the points plotted, so that approximately an equal number of points lie on each side of the line If perfect correlation exists, the line plotted will be perfectly straight, otherwise a well defined and regular curve

If the line points donward to the left, then correlation is direct and *vice-versa*, if no well defined tendency is exhibited, no correlation exists.

The graph is known as 'Galton graph'. If in this graph, both the subject and the Relative change by equal percentages, then the ratio of variation is equal to unity (one) and the line drawn through the plotted points will be a line at an angle of 45 to the horizontal and such a line represents a line of equal variation or equal proportional variation When the Relative shows a tendency to change less than the subject, the line will be at an angle less than 45 to the vertical (more than 45 to the horizontal).

If the Relative changes more proportionally than subject, the line will lie at an angle less than 45° to the horizontal.

This line is known as the regression line The nearer this regression line approaches the vertical, the slighter the degree of correlation. The larger the number of points plotted the more reliable the result will be.

The[1] Galton graph, drawn, in the annexed diagram shows that the regression line is at an angle greater than 45° that is the proportional changes in share prices are

[1] [illegible] Statistics for [illegible] p 12[illegible]

less than the proportional changes in the volume of production, or share prices fluctuate less widely than does the volume of production A numerical value is obtained by measuring the angle that the regression line makes with the vertical The ratio of the average variation of the Relative to be average variation of the Subject is represented by the tangent of the angle

Or Ratio of variation = Tan XYZ, where XY is the vertical

or $= \frac{AX}{XY}$, when A is a point where a perpendicular from a point on the vertical cuts the regression line

Equations of the lines of Regression —Mathematically equations of the lines of the regression are

(1) $y-\overline{y} = \frac{r\sigma y}{\sigma x}(x-\overline{x})$ where $\overline{x}$ and $\overline{y}$ denote the means of x and y series

Changing the origin this can be simply written as

$y = \frac{r\sigma y}{\sigma x}x$ This expresses the most probable value of y associated with a given x and it is the regression line of y on x

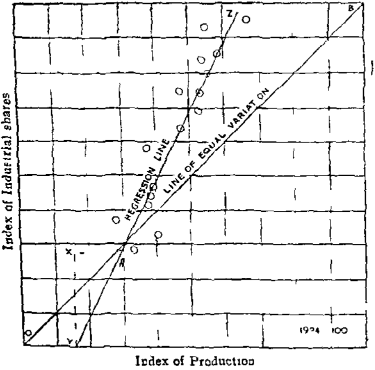
From example 1, $y = 817 \times \frac{4\ 05x}{3\ 715} = 891x$

$\frac{r\sigma_3}{\sigma x}$ is called the regression co efficient 891 is regression co efficient here.

The above regression equation gives the regression of

The Regression equation,

Galton Graph



Ratio of variation – tan ∠XYZ

$x=r\frac{\sigma_x}{\sigma_y}\,y$ *gives regression of x on y*, $r\frac{\sigma_x}{\sigma_y}$ being the co efficient of regression The regression co efficients denote the slopes of the regression lines

Significance of the co efficient of correlation —Probable and standard error of r To test whether the calculated co-efficient of correlation is significant or not the standard error or the probable error (P E) can be used standard error S E $=\frac{1-r^2}{\sqrt{n}}$

P E $=\frac{6745\,(1-r^2)}{\sqrt{n}}=$ 6745 standard error Correlation will be r ± P E

If r is less than P E correlation does not exist Correlation may be taken as good

If r is more than P E several times (at least 3)

If r is more than 6 times P E correlation is definitely good

Significance of the correlation co-efficient is also dealt with later on in Chapter XI

Exercise VIII

I — Compute the co-efficient of correlation for the following —

x	y	x	y
1	3	4	5
2	1	5	4
3	2		

Ans 6

II.—

x	*y*	*x*	*y*
7 2	17 7	3·8	17 8
11 7	21 1	5·1	18·
8·7	19 2	8 6	19 7
15 2	22 5		

Ans. ·*96.*

III.— *x*, 600 −500, −400, −200, 600,
y, −1800, 1500, 1200, 600, −1800,
700, −300, −500.
−2100, 900, 1500

Ans −*1 (perfect negative correlation).*

IV —Supply 400, 200, 700, 100, 500, 300
Demand 50, 60, 20, 70, 40, 30,

600

10

Ans −·*86.*

V —*City*	*Population (thousands)*	*Accident rate per million.*
A	10	32
B	20	20
C	30	24
D	40	36
E	50	40
F	60	28
G	70	48
H	80	44

Ans ·*71.*

VI —Find *r*, between sanitation and infant mortality for the indices of the eight cities

Sanitation 100, 86, 91, 108, 111,
Infant mortality 98, 108, 104, 98, 94,
112, 105, 87,
90, 100, 108.

VIII —Compute *r*, given.

x, 22, 27, 12, 21, 21, 27, 23, 17, 25, 1
y, 32, 27, 19, 30, 26, 26. 25, 22, 23, 5
16, 20, 37, 33, 18. 24, 22, 17, 32, [illegible]
24, 28, 29, 25, 20, 26, 17, 16, 27, 29
26, 27, 26, 21.
17, 20, 26, 17 *Ans* 55

VIII —What is the correlation co-efficient after adjustı the Probable Error in the following ? Is it significant ?

Capital in hundreds of Rupees (Subject) 10, 20, 30, 40, 50, 60, 70,
Profits in hundreds (Relative) 2, 4, 8, 5, 10, 15, 14,
80, 90, 100
20, 22, 30

Ans. '*9618* ± '*01598* Y

IX.—Draw the Galton graph from the following and show the Ratio of Variation between the following fo eight years

Year	*Subject Tense of Thousands*	*Relative in millions of Rs*
1	79	49
2	52	40
3	33	25
4	55	35
5	46	35
6	62	34
7	31	34
8	34	28

Hint.—Form the Index Numbers, by taking the av ariation $=\frac{52}{70}=$ ·74 approx. The complement of this fraction

−·74 = ·26 is called the Ratio of Regression

X —

Subject Sales, 00	*Relative Expenses 00*	*Subject Sales, 00.*	*Relative Expenses 00.*
Rs. 50	11	Rs 65	15
50	13	65	15
55	14	60	14
60	16	60	13
65	16	50	13

Find the Standard Error and the Probable Error Is the rrelation significant ?

r = ·79, P E = 08, S E = 12 Significant.

XI —

Years	(*A*) *First Differences.*	(*B*) *First Differences.*
1	−140	8
2	739	− 2
3	−620	18
4	−5486	2 3
5	1801	2 5
6	385	7.2
7	3488	−8 4
8	2576	−4·4
9	1873	−7 3
10	−5020	8 7

What is the co-efficient of concurrent Deviation ?

What would have happened if all the first differences n the two columns had the same sign ?

Ans −·*77 perfect correlation.*

XII.—

Mean annual Birth rate per 1000 of population.	*Mean annual D Rate per 1000 population*
35 3	20·8
33 5	19·4
31 4	18·9
30 5	18 7
29 3	17 7
28 2	16·0
26 3	14·7
23 6	14·3
20 1	14·4
19·9	12·2
16 7	12 1

Find r, Ans.

XIII.—Find the Regression co-efficients and lines of Regression for Q. VII.

Sol. $x = ·537 \times \frac{6·18}{5·36} y$, *and* $y = 537 \times \frac{5·36}{6·18} x$

XIV.—

Density of population per square Mile.	*(1)*	*(2)*
163	13·3	4·3
165	42·5	0·0
380	38·2	2 1
431	38·8	1·3
487	16·	1·2
440	22·4	1·2
594	15·5	3 1
710	20·2	1·6
794	28·2	3·0
2157	13·5	3 6

Find correlation between :—

Population and (1) *Ans.*

Population and (2) *A*

XV —Calculate the co efficient of correlation for the ollowing data giving the prices in ten markets of commodi ies A and B

A	61	72	73	63	84	80	66	76	74	72
B	40	52	49	43	61	58	42	58	44	45

(M A 1943) Ans 88

XVI —Find the lines of regression for the correlation able connecting Age and Marks given in the solved xample 2

(M A Aligarh 1943)

Sol —See solved example giving values of σ_x σ_y and r and then put these in the equations

XVII —Calculate the co effiaient of correlation between he prices of standard wheat and rice from the distribution iving below showing the prices in the same day in 31 markets n the Province

Prices in annas per maund of wheat

		60	64	68	72	76	*fy*
	96	2	3				5
Prices in annas per maund	102		6	2			8
	108			9	1		10
	114				5	1	6
	120					2	2
	f_x	2	9	11	6	3	31

(M A 1942) Ans 928

XVIII.—Find r from the Correlation Table.

	40—50	50—60	60—70	70 80	80—90	90—100	100—110	110—120	Totals
80—89						2	3		
70—79		8	16	20	6	6	4	3	6
60—69	7	28	26	24	8	3	3	1	1C
50—59	20	24	36	12	6	2			1C
40—49	4	8	4	2					1
30—39	6	2							
Totals	37	70	82	58	20	13	10	4	

(*M. A. Aligarh 1914*) *Ans.*

XIX.—Calculate the co-efficient of correlation for ollowing throws of 12 dice (500 in total)

Throws 2.

Throws 1.	0	1	2	3	4	5	6	7	8	9	10	11	1<
0			1	1	1								
1			1		2	3	2						
2			2	3	5	6	2	6		.			
3			5	9	8	11	16	7	6	1			
4			2	5	17	24	19	25	11	2			
5			1	5	14	25	24	24	17	4	3		
6				2	2	13	16	27	12	4	2		
7					2	7	13	22	14	5	3		
8					0	3	5	6	9	5	2		
9								2	1	2			
10									1				
11													
12													

Ans '*45* (*nearly*)

XX —(1) Define the product moment co-efficient of correlation r and prove that $-r \leqslant r \angle 1$

(2) Obtain the expression for r and the equation of the lines of regression

India Audit Accounts 1943 and M A. (Mathematics 1945)

XXI —The following marks have been obtained by a class of students in statistics (out of 100)

Paper I	80, 45, 55, 56 58, 60, 65, 68, 70, 75, 85
II	82, 56, 50, 48 60, 62, 64, 65, 70, 74, 90

Compute the co efficient of correlation for the above data Find the lines of regression and examine the relation ship

(Indian Audit & Accounts Examination 1945) $r = 92$

XXII —The following table gives the value of exports of raw cotton from India and the value of the imports of manufactured cotton goods in India for six years

Exports of raw cotton In crores of rupees	*Imports of manufactured goods*
45	50
58	53
55	58
89	65
98	76
66	58

Calculate the co efficient of correlation between the value of the exports and of the imports

Test the significance of the co efficient

Ans $r = 94$ *good (B Com. 1945)*

XXIII.—Calculate the co-efficient of correlation for short time orcillations from the following indices (1930—1944) taking a five years moving average.

x	116,	114,	111,	91,	98,	95,	92,	93,	96,	102,	107,
y	78,	84,	93,	117,	97,	102,	108,	105,	96,	77,	68,
								104,	98,	100,	108,
								77,	93,	89,	83.

Hint.—Take Moving Avarage for 5 years, takes deviations from the moving average of the corresponding indices and apply the formula. $n=11$ *Ans.* $-$·9 (*appr.*)

XXIV —Given marks as

Roll No.	1	2	3	4	5	6	7	8	9	10	11	12
Mathemetics Paper	36	56	41	46	59	46	65	31	63	41	70	36
Economics Paper	62	43	60	53	36	50	42	66	44	58	65	71

Draw a graph to show the relationship between the marks in the two subjects

Calculate the co-efficient of correlation.

$r = -$ 617 (*C. st 1945*).

✓ XXV.—Calculate the co efficient of correlation form the correlation table showing the marks obtained by 60 students in two subjects.

	5—15	15—25	25—35	35—45
0—10	1	1		.
10—20	3	6	5	1
20—30	1	8	9	2
30—40	.	3	9	3
40—50		..	4	4

Ans. ·53

XVI —Is there any relat on b tw en the series x and y given by the correlation table

x	5	10	15	20	25	30
y						
10	I	1	1	2	8	12
15		2	5	9	80	11
20	?	15	42	93	36	8
25	3	20	50	38	10	2
35	10	15	7	5	4	1

Ans Yes

XVII —Find r for the Correlation Table

x in x Rs	60—63	63—66	66—69	69—72	7?—75
100—125	2	1			
125—150		2	3	5	1
150—175		2	4	1	2
175—200			1	1	

Ans 57

CHAPTER X

MOMENTS AND NORMAL DISTRIBUTIONS

Moments play an important part as a method of comparison and in testing normality symmetry and skewness of a distribution Moments are defined about the arithmetic mean and about an arbitrary mean (or origin). Moments about arithmetic mean M are defined by the formula for rth moment as (1) for ungrouped data $\mu_r = \frac{1}{n} \Sigma (x - M)^r$ or $\frac{1}{n}\Sigma d^r$ where n is the number of items in the series $x_1 \ x_2 \ x_3 \quad x_n$ Thus the first four Moments about the mean are

$\mu_1=0$, $\mu_2=\frac{1}{n}\ \Sigma\ (x-M)^2=\frac{1}{n}\Sigma(d^2)$

which means the variance which have been studied in Dispersion, $\mu_3=\frac{1}{n}\Sigma(d_i)^3$, $\mu_4=\frac{1}{n}\Sigma d^4$. For grouped the rth moment about the mean is given by $\mu_r\ =\frac{1}{n}\Sigma f$ where n is the total number of frequencies and d the deviati of the central values from the arithmetic mean.

Moments about any provisional mean are given $V_r=\frac{1}{n}\Sigma f(D)^r$ where D denotes deviations from the visional mean, of the central values

Moments about the mean and about any provi origin are connected* as follows —

$$\mu_1=0\,,\ \ \mu_2=V_2-V_1^2\,,\ \ \mu_3=V_3-3V_1V_2+2V_1^3$$

$$\mu_4=V_4-4V_1V_3+6V_1^2V_2-3V_1^4.$$

*For Math. Proofs see Appendix.

In practice generally we need the first four momer to be calculated as follows :—Take the central values the class intervals and then deviations (D) of these fr the provisional mean (preferably the class interval hav the maximum frequency). Multiply the deviations by corresponding frequencies and add This will give Σf similarly find ΣfD^2, ΣfD^3 and ΣfD^4 to get V_2 V_3 and Put these values in μ_2, μ_3 and μ_4 to get the Mom about the Mean ain corrections (Sheppard's) may be applied for grouping nd the adjusted moments are then given as $\mu_1=0$; $\mu_2=\sigma^2=V_2-V_1^2-\frac{1}{12}i^2$, μ_3 remains unchanged and $\mu_4 V_4-4V_1V_3+6V_1^2V_2-3V_1^4-\frac{1}{2}\mu_2 i^2+\frac{7}{240}i^4$ where i denotes class interval

Normal Distributions—In dealing with graphs on requency distribution, a smoothed curve, a bell shaped ur e has been drawn This smoothed curve may be a continuous and perfectly symmetrical curve known as the Normal curve stretching to infinity at both ends (Figure next Chapter) and it is the curve representing Normal distributions.

To determine whether a given distribution is Normal, we have to determine some other statistical parameters known as α, β, γ define as —

$$\alpha_1=\frac{\mu_1}{\sigma},\ \alpha_2=\frac{\mu_2}{\sigma^4}\ =1,\quad \alpha_3=\frac{\mu_3}{\sigma^3}\ =\sqrt{\beta_1}$$

$$=\gamma_1=\frac{\mu_3}{\mu_2^{\frac{3}{2}}}$$

$$\alpha_4=\frac{\mu_4}{\sigma^4}=\frac{\mu_4}{\mu_2^2}=\beta_2=\gamma_2+3\quad \beta_1 \text{ and } \beta_2$$

are the measure of Symmetry and Normality. If $\beta_1=0$ the distribution is symmetrical if α_4 or $\beta_2=3$, the distribution is Normal. The quantity α_4 measures a characteristic called 'Kurtosis' i e, flatness of the curve, α_4-3 is called the Excess over the Normal distribution If $\alpha_4 <_3$ the curves is said to be platykurtic (flat-topped and short-ailed) if greater than 3, then if it is said to be leptokurtic. (Peaked more sharply and long-tailed).

For a normal curve $\beta_1=0$, $\beta_2=3$ & Excess, $E=0$.

Skewness can also be measured in terms of β_1 and β For a large class of curves to which the moderately ' is a close approximation, the skewness is given by

$$\frac{\sqrt{\beta_1}\quad(\beta_2+3)}{2(5\ \beta_2-6\ \beta_1-9)}$$

Example.—Calculate the four Moments for the follow distribution of wages after applying Sheppard's corrections

Weekly earnings,	*Men*		
x Rs.	*f*	*D form*	*f D*
		a Mean 10	
5	1	−5	− 5
6	2	−4	− 8
7	5	−3	−15
8	10	−2	−20
9	20	−1	−20
10 —	51	0	0
11	22	1	22
12	11	2	22
13	5	3	15
14	3	4	12
15	1	5	5
	131		8

Performing the calculations, we shall have

$$V_1=\frac{\Sigma\ f\ D}{n}=\frac{8}{131}=06,$$

$$V_2=\frac{\Sigma\ f\ D^2}{n}=\frac{346}{131}=2{\cdot}64,$$

$$V_3 = \frac{\Sigma f D^3}{n} = \frac{74}{131} = 56,$$

$$V_4 = \frac{\Sigma f D^4}{n} = \frac{3718}{131} = 28\ 38$$

Hence, using the formula for μ_1, μ_3,/and μ_4 in terms V_1 V_2 after applying Sheppard's corrections, we have after alculation, $\mu_2 = V_2 - V_1^2 - \frac{1}{12} = 2\ 55$

$$\mu_3 = 57 - (3 \times 2\ 66 \times\ 06) + 2 \times (\ 06)^3 = 085$$

$$\mu_4 = 28{\cdot}4 - (4 \times\ 06 \times\ 56) + 6 \times (\ 06)^2 \times 2\ 64$$

$$-3(\ 06)^4 - \tfrac{1}{2}(2\ 56) + {\cdot}029 = 27 \text{ nearly}$$

If we want to test the symmetry and normality, then find $_1$ and β_2

Now, $\beta_1 = \frac{({\cdot}078)^2}{(2\ 55)^3} = 00036$ (approx)

$\beta_2 = \frac{27}{(2\ 55)^2} = 4$ (approx) $= \frac{\mu_4}{\mu_2^2}$

Here $\beta_2 > 3$, so that the distribution is leptokurtic and not normal As β_1 is very small, symmetry exists

Exercise IX

a I.—Find the first four moments about the mean for the lata in Q I and II.

x	1,	2,	3	4,	5,	6,	7	8,	9
'	1,	6,	13,	25,	30,	22,	9	5,	2

Aligarh University M.A. 1942)

(*Ans* $\mu_2 = 2\ 478$, $\mu_3 = {\cdot}679$, $\mu_4 = 18{\cdot}35$

X.

II.—

30—40,	40—50,	50—60,	60—70,	70—80,	80—90,	90—100
2	3	11	20	32	25	7

Find also β_1 & β_2

Ans 172, ~1320, 94096,
β_1= 34, β_2=3 17.

III —Compute the first four moments about an arbitrary origin from the following frequency distribution of heights in inches of adult Irishmen.

Height	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73
Adults	1	0	2	2	7	15	33	58	73	62	40	25	15	10	3

(Punjab University M A 1942)
Ans- ·341, 4 821, 4 468, 81·61

IV —Data from a fisheries investigation, x being the numbers of tail rays in 703 flounders Find a_4 and test for Normality

x	47,	48,	49,	50,	51,	52,	53,	54,	55,	56,	57,	58,	59	60,	61
f	5,	2,	13,	23,	58,	96,	134,	127,	111,	74,	37,	16,	4,	2,	1

(M.A 1942) 3·3, leptokurtic.

V —Calculate the first four moments about Mean of the distribution of Weights given by the following data after applying Sheppard's corrections.

Weights Seers, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77,

Men 2, 4, 14, 41, 83, 169, 394, 669, 990, 1223, 1329, 1230, 1063, 646, 392, 202, 79, 32, 16, 5, 2.

Ans. 6·533, —·208, 134 409.

VI.—Find β_1 and β_2, skewness and Kurtosis for Q. V.

Ans skewness = 006 , leptokurtic β_1= *00015 &* β_2=*3 14*

VII —Given x in Annas 156, 159, 162, 165, 168,
Men · 3, 9, 26, 53, 89,

171, 174, 177, 180, 183, 186, 189, 192, 195, 198
146, 188, 181, 125, 92, 60, 22, 4, 1, 1

Find the moments about the Mean and α_3, α_4.

Ans. 43 35, −9 77 , 5508·56 ;

α_3= − 033 · α_4=2·92; (approximately Normal)

VIII.—Find the standard deviation, adjusted β_1 and β_2 (after Sheppard's correction) for Q VII and test for normality σ=6·5 ; β_1= 0012 β_2=2 93 (approx. Normal)

IX —Find μ_2, μ_3, μ_4, β_1, β_2 after Sheppard's correction for

x 3, 8, 13, 18, 23, 28, 33, 38, 43, 48, 53, 58, 53
f 5, 9, 28, 49, 58, 82, 87 79, 50, 37, 21, 613

Ans 5 34, 0292, 76·1143.
β_1 & β_2= 00056, 2 663

X.—Find moments after corrections for

x 59, 61. 63, 65 67· 69 71
f 1, 29, 48, 131, 102, 40, 13

Ans. 4 7, −·89, 82 85.

XI.—Derive the expressions for moments about the Mean terms of the moments about any arbitrary origin

Calculate the moments for the following —

Marks	20—30,	30—40,	40—50,	50—60,	60—70,	70—8	80—90,	90—1
Frequency	6	28	96	75	56	30	8	1

M A (Math

Ans V_1, = *— 087, 175 3, 313 3 & 82333*

XII —Calculate the first four moments, β_1 & β_2 and for symetry and normality

XII —

Weekly earnings	15,	16,	17,	18,	19,	20,	21,
Labourers	8,	10,	15,	20,	25,	30,	40,

(*C St 1945*) *Ans*

CHAPTER XI

ELEMENTS OF PROBABILITY, SAMPLING, *TEST OF SIGNIFICANCE AND ANALYSIS* OF VARIANCE

The theory of Probability plays a very important p not only in Statistics but in all sciences Here we sl explain its meaning very briefly

If an event can happen in m ways and fails in n w and each of these $(m+n)$ ways are equally likely to the probability of the happening of the event is $\frac{m}{m+n}$ and that of its failling is $-\frac{n}{m+n}=q$

The sum of the probability of the success p failure q is $=1$

When a coin is tossed a head or a tail is equally ely to fall therefore $p=\frac{1}{2}$ and $q=\frac{1}{2}$ The probability of awing an ace from a pack of 52 cards is $\frac{4}{52}$

Events may be independent, dependent and mutually clusive An event E is said to be independent of another (1) ent F when the actual happening of F does not influence any degree the probability of the happening of E If e probability of the happening of E is dependent on, or (2) fluenced by the previous happening of F then E is said to dependent on F

Two events E and F are said to be mutually exclusive (3) hen through the occurence of one of them, say F, the her event E cannot take place or *Vice Versa*

Theorem of Addition of Probabilities —When an event may happen in any one of the n different and mutually clusive ways, E_1, E_2 with probabilities p_1, p_2 p_n, n the probability for the happening of the event E is ual to the sum of the probabilities p_1+p_2+ p_n

Theorem of multiplication of probabilities *—The probility, p, for the simultaneous or consecutive appearance several mutually exclusive events is equal to the product $\times p_2 \times p_3 \times$ p_n The theorem is called, Theorem on compound probability' A card is drawn from a packet and placed by a joker, then a second card is drawn The obability that both cards are aces is $p=p_1\times p_2=\frac{4}{52}\times\frac{3}{52}$. no replacement is made, (Dependents Events), and a ond card is drawn, then $p=\frac{4}{52}\times\frac{3}{51}$.

If two coins are tossed, there are altogether 4 ways of ir falling

H H H T, T H T T

$p = \frac{1}{4} \qquad \frac{1}{2} \qquad \frac{1}{4}$, Sum$=1$

If n coins be *tossed the frequency* distribution of the respective chances of n $n-1$, $n-2$, 3, 2 1, 0 heads is given by $(\frac{1}{2}+\frac{1}{2})^n$

In general if p and q represent the probabilities of success and failure for a single event $(p+q=1)$ the frequency distribution of the Chances of n, $n-1$, $n-2$, 2, 1 0, successes in the compound event is given by t successive terms of the binomial expansion

$$(p+q)^n = p^n + n\,p^{n-1}q + {}^{n}\frac{c(n-1)}{2}\,p^{n-}\ q^2 + \qquad + q^n$$

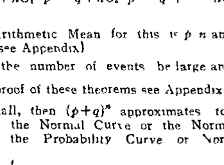
$$\text{Or } p^n + nC_1\ p^{n-1}q + nC_2\ p^{n-2}q^2 + \qquad nC^r\ p^{n-r}\ q^r + + q^n$$

The Arithmetic Mean for this is $p\ n$ and $\sigma = \sqrt{n\,p\,q}$ (For proof see Appendix)

If n the number of events be large and neither p n q very small, then $(p+q)^n$ approximates to the regu curve, i e the Normal Curve or the Normal Frequenc Curve or the Probability Curve or Normal Curve o Error

*For proof of these theorems see Appendix

NORMAL CURVE



Normal Curve—In the Normal Curve as shown in the figure the origin is taken at the centre O, the variable is measured along the x axis and its frequency along the y axis The chances of a deviation from the centre according to the Normal Curve are given as,

Deviation lying between	*Chance*
$5\,\sigma$ and $-\,5\,\sigma$	383
$6745\,\sigma$ and $-\,6745\,\sigma$	5
σ and $-\sigma$	682
$2\,\sigma$ and $-2\,\sigma$	954

The quantity 6745σ is said to be the Probable error

Thus for a variable there is a chance of 682 in 1000 that a deviation from the Mean will not exceed the standard deviation σ and a chance of 318 in 1000 or nearly 1 in 3 times that it will

Similarly a deviation greater than 2σ will occur 46 in 1000 or nearly 1 in 19 or 20 Following this a probabili y of 19 to 1 against an occurrence is generally taken, as the criterian of significance though some people use 99 to 1 depending on the nature of the Variables Tables given above give the probability of obtaining the deviations of any size Such a table is known as a Table of Probability Integral and may be found in Pearson's Tables for Statisticians and Biometricians

The equation of the Normal Curve can be written as

$$y = \frac{N}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2\sigma^2}} \quad \text{or} = \frac{N}{2\;506\,\sigma}\, 2\;718^{\frac{-x^2}{2\sigma^2}}$$

N being the total number of frequencies.

To fit the Normal curve, assign appropriate val (multiples of σ) to x We can obtain from this eq the corresponding value for y, or the Ordinates for the curv After securing some ordinates, the curve can be tr fitted to the given data Putting $x=0$, we get the maxim ordinate, to be erected at the Arithmetic Mean.

If the magnitude of the class intervals (i) is cons then the equation can be written with Ni instead of In practice straightforward normal distributions of phy objects are rare except in certain branches of biolog science.

A'Priori and Empirical Probabilities.—The probab described above is of the a'priori type *i e.* chances are de from consideration of the possible combinations of involved In the majority of cases in practical life, factors at work are not definitely known. So Empirical bability is used which is based upon actual observation experiment. Let m trials be made of which s represent ses and f, failures. The best estimate of the chance of event happening is

$p=\frac{s}{m}$, and of non-occurrence is $\frac{f}{m}$. Empirical bability depends on the number of trials, if the number is large, the estimate wil also be very accurate. According to life table, out of every 100,000 persons living at age 10, 82 survive to age 40, of whom 820 die during the year, 81,455 survive till 41, therefore the probabili of a life aged 40, dying within the year w be $\frac{820}{82275}=01$ and the chance of his surviving t

year $=\frac{81455}{822775}=·99.$

Sampling—For a comprehensive and bulky data, much time, energy and money would be needed for its statistical analysis From the big mass of data a part or parts may be taken to represent the whole This small mass selected from the big mass is said to be Representative Data' or a sample The process of such selection is sampling The whole data is called the population or universe from which samples are taken

Sampling may be (1) deliberate or purposive with definite objects in view (2) Random with no definite purpose (3) Stratified i e to segregate a heterogeneous universe into homogeneous sub groups and to draw from each sub group a sample at random If (1) and (2) combine, it will be mixed sampling Random selection consists in picking up at random from a big mass, such a few examples, as can sufficiently represent the whole population The examples thus selected are studied intensively As the deliberate selection is likely to be prejudiced so random selection is preferred Surveys formed are known as sample surveys

General laws of Statistical Induction (1) The law of Statistical Regularity which lays down that a group of objects chosen at random, from a larger group tends to possess the characteristics of the whole universe The sample should not be too small as it may be biassed or it may not be representative

Every item in the population must stand an equal chance of being included Lots may be drawn for randomness and Tippett's Random Numbers may also be used (Random numbers are also given in Fisher and Yates Tables) The larger the sample the more reliable are its indications

(2) The law of Inertia of Large Numbers It fell from (1) and according to it, large aggregates are relat more stable then the small on s If the numbers involve are of great magnitude, the total change will likely almost insignificant For instance, while the production wheat may differ from place to place, owing to the scarcity rain, the visit of floods or some other cause, the total producti of the World as a whole remain fairly constant

Both the laws are based on experience and the insura principles are based on these The theory of sampling based on probability

Sampling fluctuations or Errors of sampling No sample can afford a perfect representation of t universe from which it is drawn Inspite of precautions secure randomness, variations occur, due to the elements chance present in the selection Such variations are known sampling Errors or fluctuations The reliability of the sampl depend upon their probable magnitudes

Measures of Reliability or Tests of significance In general the results of sample inquires will sho differences that cannot be assigned to any definite caus Every sample will have its peculiarities in the for n 1.frequency distribution and in the magnitude of 2its averag 3standard deviation and 4skewness

|* These differences are the fluctuations and it is the a of the theory of sampling based on the theory of Probabilit to supply tests with the help of which it can be determ whether any given fluctuation is statistically significant or n

We shell deal in this chapter with the significance

Mean, Differences between means, significance of standard deviation and of the correlation co-efficient r

It will be found that a large number of experiments show many different values of the mean, each one departing more or less from the true mean of the entire universe If the standard deviation of the whole population is $\overline{\sigma}$ and we take a large number of random samples of n observations, then the means of the samples will be distributed with a standard deviation $\frac{\overline{\sigma}}{\sqrt{n}}$ If the universe is normally distributed, the means also will be normally distributed If the distribution of the universe is not normal the distribution of the Means of samples still tends to be normal provided the size of the samples is sufficiently large, but in cases of small samples the distribution of the means is not normal

The standard deviation of the entire population (sometimes called parent population) is not generally known, so we have to take the standard deviation of an observed sample as an estimate of it The standard deviation of the sampling distribution is then estimated from the standard deviation of a single sample This estimated value i e, the standard deviation of the mean of a random sample, is called the standard error of the mean and is given by

$$\sigma_M = \frac{\sigma}{\sqrt{n}},$$

where σ is the stanbard deviation of the sample and n the number of observations in it. Probable Error of the Mean $= \frac{\cdot6745\,\sigma}{\sqrt{n}}$ (though the probable error is not much used in p actice)= 6745 standard error (nearly $\frac{2}{3}$ S E) T best estimate of the mean of the population is M $\pm \frac{\sigma}{\sqrt{n}}$ The larger n the smaller the standard error when n is sufficiently large, the standard error is almos negligible In examining the significance of the Mean of n items x_1 x_2 x_n for small sample the standarc deviation is to be found

by $\sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}$, where $n-1$ represents the number degrees of freedom for calculation σ After obtaining x and σ, in this way, we obtain the 't' statistics which is essentially the ratio of the mean to its standard error or

$$t = \bar{x} \div \frac{\sigma}{\sqrt{n}} = \frac{x\sqrt{n}}{\sigma}$$

This 't distribution is due to 'Student

The following table (from Fisher and Yates tables III) gives values of 't corresponding to different values of N ne number of the degrees freedom (one less than the total number of observations)

If the calculated value of t' is greater than that given in the table for appropriate value of N, then the mean is significantly different from zero, otherwise not

Table values of 't' corresponding to a probability P = 05 [levels of significance]

N	*t*	*N*	*t*	*N*	*t*
1	12 706	13	2 160	25	2 060
2	4 303	14	2 145	26	2 056
3	3 182	15	2 131	27	2 052
4	2 776	16	2 120	28	2 048
5	2 571	17	2 110	29	2 045
6	2 447	18	2 101	30	2 052
7	2 365	19	2 093	40	2 021
8	2 306	20	2 086	60	2 000
9	2 262	21	2 080	$N = \propto$	$t = 1\ 96$
10	2 228	22	2 074		
11	2 201	23	2 069		
12	2 179	24	2 064		

When the number of observations is large, calculate $\frac{\sigma}{\sqrt{n}}$ in the usual way

Example —Eleven school boys were given a test in Geometry they were given a month's further tuition and a second test of equal difficulty was held at the end of it Do the marks give evidence that the students have benefited by the extra coaching ?

Boys	*Marks*	*Marks*	*Differences*
	1st Test	*2nd Test*	*x*
1	23	24	+1
2	20	19	−1
3	19	22	+3
4	21	18	−3
5	18	20	2
6	20	22	2
7	18	20	2
8	17	20	3
9	23	23	0
10	16	20	4
11	19	17	−2 +11

The problem is ' Is the mean of the differences between the marks of the two tests significantly different from zero ?

A Mean= $\frac{11}{11}$ =1 and standard deviation (of Drob

$$\sigma = \sqrt{\frac{50}{11-1}} = \sqrt{5}$$ ✓

Hence standard error of the mean $\sigma_M = \frac{\sigma}{\sqrt{n}}$

$\sigma_M =$ $\sqrt{5} \times \frac{1}{\sqrt{11}}$ applying 't test

$$t = 1 \times \frac{\sqrt{11}}{\sqrt{5}} = 1\ 3$$

The calculated value is less than tha value of t 175) for $N = 11 - 1 = 10$ in the table Hence the mean of x is sign ncan y diff rent from zero and the marks are not enough to prove th adv intage of extra coaching

S gnificance of the d ff rence between two means statis ical language the problem may be expressed as Is diff-rence between the means su h that they might been drawn from the same un verse by random or are they drawn from two diff rent universes or p lations? For a large numb r of observations in the samples the standard error of the difference means is g ven by $\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ where σ_1 and σ_2 standard deviations of the two independent (as given by two diff rent authorities) and uncoreelated ables n_1 and n_2 are the number of observations in sample

erence
mean

If the d ff-rence between the two means is than twi e its standard error, then the means signifi a ly d fferent

Given the following

		Marks	
Mean	-	$\overline{x_1}=130$	$x_2=127$
S D		$\sigma_1= 14,$	$\sigma_2= 12$
of boys in Class I & II		$n_1= 84,$	$n_2= 60$

The problem is to find whether themean test score of Class I is significantly greater than that of the Class II erence between the Means$=130-127=3$

of difference $-\quad \sqrt{\frac{196}{84}+\frac{144}{60}}=2\ 2$ nearly

The difference between the Mean is less than twice standard error bence the mean test score Class I not significantly greater than of the Class II For ll samples t test has to be applied —If x_1 x_2 are two s of observations with means $\overline{x_1}$ and $\overline{x_2}$ respectively find expression

$$\frac{\Sigma(x_1-\overline{x_1})^2+\Sigma\ x_2-x_2\ )^2}{(n_1-1)+(n_2-1)} -\text{S (say)}$$

Where n_1 & n_2 are the number of observations or uencies The value of t is given by

$$=\frac{\underset{1}{x}-\overline{\underset{2}{x}}}{\text{S}\times\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

For the t table the degrees of freedom will be

$\text{N}=n_1-1+n_2-1$

The significance is tested in the same way as for mean

Standard Errors for, Median, for standard ' tion, and Mean deviation, etc., are the following —

$$\sigma_{\text{Median}} = 1\,253 \frac{\sigma}{\sqrt{n}},\ \sigma_{\text{S deviation}} \text{ or } \sigma_{\sigma} = \frac{\sigma}{\sqrt{2n}}$$

$$\sigma_{\text{M Dev}} = 6038 \frac{\sigma}{\sqrt{n}} \text{ Probable error} = {}^{\cdot}6745 \text{ S E}$$

$\sigma_{\gamma_1} = \sqrt{\frac{6}{n}}$, $\sigma_{\gamma_2} = \sqrt{\frac{24}{n}}$. If γ_1 and γ_2 are both than twice (at least) their S E. then the distrib not significantly different from the normal form In when n is large, the error becomes smaller and smaller

Significance of the Co-efficient of Correlation r

Standard error of $r = \frac{1-r^2}{\sqrt{n}}$. This is approx true when n is large In such cases the correlation taken as differing significantly from zero, if r is than twice (at least preferably thrice) its standard The standard error is generally applied when $n = 5$(more In small samples, however, the significance of r be determined with the help of 't' tests, which is by $t = \frac{\sqrt{n-2}}{\sqrt{1-r^2}} r$ if the value of t is larger than the given in the table for 't' for $n-1$ degrees of freedom, th significantly larger than zerro Moreover Fisher and Y (Table VI) have given tables for the values of r for P etc., and $N = n-2$, The calculated value of r may sim be compared with the values of r in the table for signi and degree of association

The 'z' *test* [illegible]

orrelation Letween two variables is different in two ent samples, Fisher's 'z' test method is used. rding to this method r is transformed into 'z' such that

$$z=\tfrac{1}{2}\log_e \left\{\frac{1+r}{1-r}\right\}$$

The values of z corresponding to the values of r are in Fisher and Yates tables

he standard error of z is $\dfrac{1}{\sqrt{n-3}}$ and the standard error

difference between two z's is

$$\sqrt{\frac{1}{n_1-3}+\frac{1}{n_2-3}}$$

n_1 and n_2 are the numbers of pairs in the two es If the difference between the two z's is greater twice its standard error, then the difference is significant

ysis of Variance

The analysis of variance is a useful method in scientific ies especially in Agriculture and Biology It may iefly explained as follows It is known that the ice of a variable is obtained from the sum of the s of the deviations of items from the general Mean um of squares can be split up into two portions Suppose we have yields per acre of 6 plots of wheat, three of plots are of variety a and three of variety b.

a	..	30	32	22
b	...	20	18	16

1) The total sum of squares is separated into one

due to variation *between* the varieties Let the mean all observations $x_1, x_2 -$ ·be denoted by $\bar{x}$, which in case is $\frac{138}{6}=23$ The mean of a, $\bar{x}_a$ is $\frac{84}{3}=28$ and of is $\frac{54}{3}=18$

(2) Find $\sum_1^6 (x-x)^2$ which in his case is $(30-/$ $+(32-23)^2+(22-23)^2+(20-23)^2+(18-23)^2+(16-23)^2$ $=49+81+1+9+25+49=214$

(3) Find the sum of squares for within the varie for (a), it is $(30-28)^2+(32-28)^2+(22-28)^2=4+16$ $=56$ for (b) it is $(20-18)^2+(18-18)^2+(16-18)^2=4+0$ $=8$

The total sum is $56+8=64$ i e we have fr $\sum_1^3 (x-\bar{x}_a)$ and $\Sigma(x-\bar{x}_b{}^2$

(4) Find the sum of squares for *between* the variet This is given by $3\times[(28-23)^2+(18-23)^2]=3$ [25 $=150$ (We obtain deviations of the means of a and from the general mean square and then sum The w sum is multiplied by 3 because each value represents mean of 3 plots) i e we have found $3[\Sigma(\bar{x}_a-x)+$ $\Sigma(\bar{x}_b-\bar{x})^2]$ or $3\;\Sigma(\bar{x}_i-x)^2$ where $\bar{x}_i$ represents the mean of group

(5) Adding (3) and (4) we obtain the sum as $64+$ $=214$, which is the same as in (2)

The sum of squares can a ways be divided in this way nto two components or parts.

In general if there are k groups and n simple observa-ions in each group, then

$$\sum^{nk} (x-\bar{x})^2 = \sum^{k} \sum_{i}^{n} (x-\bar{x}_t)^2 + n\sum_{t}^{k} (\bar{x}-\bar{x}_t)^2$$

In the above $n=3$ nd $k=2$.

The degrees of freedom corresponding to the sum of quares are given for,

Total	Within	Between
$(nk-1)$	$k(n-1)$	$k-1$

nd the variance is obtained by dividing the sum of squares y the degrees of freedom The analysis of variance is set p in the following table —

Source of Variation	*Sum of Squares*	*Degrees of Freedom*	*Variance*
Vithin the Varieties	64	4	16
etween the Varieties	150	1	150
	214	5	

The variance for between the varieties is very high as ompared to that for within varieties Generally if the ariances are significantly different, there is some specific ause of variation. To test the significance find the ratio $\frac{v_2}{v_1} = F$ where v_1 denotes the variance for ' within the arieties (known also as Error Variance) v_2 denotes the ariance for ' between the varieties.'

Tables have been constructed by Snedecor for F, for n_1 and n_2 degrees of freedom, n_1 being the degrees of freedom or ' $\overset{2}{\text{within}}$ ' and n for ' $\overset{1}{\text{between}}$ ' the varieties.

If the calculated value of F is greater than the value of F in the tables for n_1 and n_2 degrees of freedom, then the differences between the varieties is significant

Logarithms also can be used Fisher has given tables for z for the testing the significance where

$$z = \frac{1}{2} \log e \text{ F}$$

Exercise X.

1 Find the probability of throwing 9 with three dice

Ans. $\frac{25}{216}$.

2 In the Binomial distribution $(p+q)^n$ find the Mean and standard deviation if $p = 3$ and $n = 20$ What is q ?

Ans 6 2 04, 7

3 A bag contains 5 white and 7 black balls If 2 balls are drawn, what is the probability that one is white and the other black ?

(M A 1943) Ans $\frac{5c_1 \times 7c_1}{12c_2} = \frac{35}{66}$

4 In solved Ex 1, let x be 4 1, 0, 3, 4 -4, 2, -2, 1, 1, 1 Determine the significance

Ans. No

5 One purse contains 1 sovereign and 3 shillings, a second one contains 2 sovereigns and 4 shillings, and a third one contains 3 sovereigns and 1 shilling If a c is taken out of the purses selected at random, find the chance that it is a sovereign

(M [illegible]) *4*

6.	*Number of persons examined*	*Mean height*	σ
(1)	600	67 5"	2 55
(2)	1300	68 6"	2 5

Find whether the persons in (2) are significantly taller than those from (1)

Ans. Yes.

7 Apply 't' test to find whether correlation is significant, if $r = \cdot 6$ and $n = 38$

Ans. Yes

8 Apply 'z' test to determine the significance between two correlations given by

$r_1 = \cdot 472$, $r_2 = \cdot 377$, $n_1 = 42$, $n_2 = 39$

Ans No

9 Fit a Norma curve for a frequeny distribution whose class interval is 10, σ is 21, Mean = 80 6, $n = 300$

Hint —Take x as $\pm \frac{1}{5}\sigma \pm \frac{2}{5}\sigma$ to find ordinates Trace the curve

10 Set up a table of analysis of variance, for

Plots	*Varities* (1)	(2)	(3)	(4)
a	140	145	150	160
b	100	110	120	125
c	200	180	160	150

11. Give two series

x_1	21·3,	20 8,	23 7,	24 3,
x_2	20·2,	16·9,	18·2,	16·7

Is there significant difference between the two means

12 Calculate the sampling error of the mean in Q 24 Ex I

(C St 1945)

Ans 0172.

13 Set up a table of analyses of variance and find F, for the varieties of gram in the following plots

1	27 6	32 4	23 4
2	19 2	18 6	16 5

Ans 12 8

14 Set up a table of analysis of variance for yields of four strains of wheat planted in five randomised blocks

Strains	Blocks			
a	30	35	38	36
b	22	40	42	44
c	34	45	50	36
d	30	50	45	38

15 A bag contains 6 white and 9 black balls Two drawings of 4 balls are made such that, (*a*) the balls are replaced before the second trial (*b*) the balls are not replaced before the second trial Find the probability that the first drawing will give 4 white and the second 4 black balls in each case

Ans $\frac{6}{65}$, $\frac{21}{55}$

(Indian Audit and Accts 1945)

16 A can hit a target 3 times in 5 shots, B 2 times in 5 shots, C 3 times in 4 shots They fire a volley Wheat is the probability that 2 shots hit?

17 A bag contains K sin ilar balls A part cf or a balls are drawn

What is the probability of drawing (1) an even numb (2) odd number of balls (M A Aligarh 1943)

$$\begin{array}{lcc} & k-1 & k-1 \\ Ans & \underline{2-1} & 2 \\ & k & k \\ & 2-1 & 2-1 \end{array}$$

18. A and B throw with one dice for a prize of Rs 1l which is to be one by the player who first throws 6 If A has the first throw what are their respective expectations ?

(*Hyderabad B Sc 1945*) *Ans Rs 6 & 5*

19 It is 8 5 against a person who is now 40 years old living till he is 70 and 4 3 against person now 50 living till he is 80 Find the probability that one at least of these persons will be alivo 30 years hence

Ans $\frac{59}{91}$ (*Hyderabad B A 1945*)

CHAPTER XII

INTERPOLATION AND GRADUATION

In Chapter IV Interpolation was explained graphically In this chapter we shall deal with formulæ for interpolation

(1) Newton's formula for equidistant spaces (equal gaps or equal intervals)

Let x be the independent variable or the argument, y or $f(x)$ the corresponding value for x or the function of x

Given	x	y or $f(x)$
	a	$f(a)$
	$a+w$	$f(a+w)$
	$a+2w$	$f(a+2w)$
	$a+3w$	$f(a+3w)$
		...

To find the value of y or $f(x)$ for a value of x lying, somewhere between a and the last item in x say for $a+xw$ Newton's formula gives

$$f(a+xw)=f(a)+x\Delta f(a)$$

$$+\frac{x(x-1)}{2}\Delta^2 f(a)+\frac{x(x-1)(x-2)}{3!=3\times 2}\Delta^3 f(a)$$

$$+\frac{x(x-1)(x-2)(x-3)}{4!=4\times 3\times 2=24}\Delta^4 f(a)$$

$$+\frac{x(x-1(x-2)(x-3)(x-4)}{5!=120}\Delta^5 f(a)+$$

Where $\Delta f(a)=f(a+w)-f(a)$, known as the firstdifference of $f(a)$ or Difference of first order'

$\Delta^2 f(a) = \Delta f(a+w) - \Delta f(a)$ known as the difference of $f(a)$ or difference of second order

$\Delta^3 f(a) = \Delta^2 f(a+w) - \Delta^2 f(a)$ known as difference o third order

$$\Delta f(a+w) = f(a+2w) - f(a+w), \quad \Delta^2 f(a+w$$
$$= \Delta f(a+2w) - \Delta f(a+w) \text{ and so on.}$$

Example 1 —The population oi India in the followin four censuses isgiven in millions, to find the population fo 1926

x	$f(x)$	Δ	Δ^2	Δ^3
1901, a	294			
		21		
1911, $a+w$	315		-17	
		4		47
1921, $a+2w$	319		30	
		34		
1931, $a+3w$	353			

(*M.A. Aligarh 1943*)

To find the population for 19'6, put $a+xw=1926$, since a is 1901 and w is 10, . $x=\frac{5}{2}$

The first difference $\Delta f(a)$ is given by subtracting the upper value from the lower value or entry in $f(x)$ Second differences Δ^2 are obtained by subtracting the upper value in column Δ from the lower value and so on

These differences are placed in between in column Δ, Δ^2, Δ^3 as shown in the table

In this way we proceed further When this table known as Difference Table has been formed, we apply

Newton's formula, taking the topmost values for $\Delta f(a)$, $\Delta^2 f(a)$ etc. (known as the leading diagonal)

$$\therefore\ f(1926)=294+\tfrac{5}{2}\times 21+\frac{\frac{5}{2}(\frac{5}{2}-1)}{2}\times(-17)+$$

$$\frac{5}{2}\,\frac{(\frac{5}{2}-1)(\frac{5}{2}-2)}{6}\times 47=294+\frac{1695}{48}=329\frac{5}{16}\text{ millions.}$$

Newton's formula is also written in the form

$$u_x=u_0+x\,\Delta u_0+\frac{x(x-1)}{2}\,\Delta^2 u_0+\frac{x(x-1)(x-2)}{6}\,\Delta^3 u_0+\ \ldots$$

u_0 stands for $f(a)$ and u_x for the interpolated value. The series will terminate after some differences. The results obtained will in general be approximate depending largely on the nature of data and circumstances governing the data being normal

2. Lagrange's formula for unequal intervals Given the following

x	$f(x)$
a	$f(a)$
b	$f(b)$
c	$f(c)$
d	$f(d)$
e	$f(e)$
	...
...	...
	...
...	...

where a, b, c, d, ...differ by unequal gaps or intervals. To find the value for any other x, the Lagrange's formula is used, which is stated as

$$f(x)=f(a)\times\frac{(x-b)(x-c)(x-d)(x-e)}{(a-b)(a-c)(a-d)\ \ldots\ .}+$$

$$f(b) \times \frac{(x-)(x-c)\ x-d)}{(b-a)(b-c)\ b-d)}$$

$$+f(c) \times \frac{(x-a)(x-b)(x-d)}{(c-a)(c-b)(c-d)} \quad +$$

$$f(d \times \frac{(x-a\ (x-b)(x-c\ \ x-e}{(d-a)(d-b)(d-c)(d-e)} + \quad -$$

Example 2 —Given

x	$f(x)$
14	68 7
17	64
31	44
35	59 1

(*M A Purjab 1942*)

To find the value for $x=27$

In Lagrange's formula put $x=\,{}_{\llcorner}7$, $a=14$, $b=17$, $c=31$ and $d=35$

$$f(27)=68\ 7 \times \frac{(27-17)(?7-31)(27-35)}{(14-17)\ 14-31)(14-35)}$$

$$+64 \times \frac{(27-14)(\llcorner 7-31)(27-35)}{(17-14)(17-31)(17-35)}$$

$$+44 \times \frac{(\llcorner 7-14)(27-17)(27-35)}{(31-14)\ 31-17)(31-35)}$$

$$+39\ 1 \times \frac{(27-14)(27-17)(27-31)}{(35-14)\ 35-17)(35-31)}$$

$$=49\ 3 \text{ nearly}$$

, Lagrange's formula is also written as —

$$u_x = u_0 \times \frac{(x-b)(x-c)}{(a-b)(a-c)}$$

$$+u_1 \times \frac{(x-a)(x-c)}{(b-a)\ b-c)\cdot\ -} +$$

Central Difference Formulae —The following we known formulae are also used for interpolation, the ment $x=a$, being taken in the middle and the differenc table being in the form

argument x	$f(x)$ or *entry*
$a-2w$	$f(a-2w)$
$a-w$	$f(a-w)$
a	$f(a)$
$a+w$	$f(a+w)$
$a+2w$	$f(a+2w)$
...	...

(1) Gauss formula

$$f(a+xw)=f(a)+x\,\Delta f(a)+\frac{x(x-1)}{2}\Delta^2 f(a-w)$$
$$+\frac{(x+1)\,x\,(x-1)}{3!}\Delta^3 f(a-w)$$
$$+\frac{(x+1)\,x\,(x-1)(x+2)}{4!}\Delta^4 f(a-2w)+$$

(2) Stirling formula

$$f(a+xw)=f(a)+x\,\frac{\Delta f(a)+\Delta f(a-w}{2}$$
$$+\frac{x^2}{2!}\Delta^2 f(a-w+\frac{x(x^2-1^2)}{3!}\times$$
$$\frac{\Delta^3 f(a-w)+\Delta^3 f(a-2w)}{2}$$
$$+\frac{x^2}{4!}(x^2-1^2)\,\Delta^4 f(a-2w)+$$

(3) Bessel's formula

$$f(a+xw)=\tfrac{1}{2}\{f(a)+f(a+w)\}+(x-\tfrac{1}{2})\,\Delta f(a)$$
$$+\frac{x(x-1)}{2!}\tfrac{1}{2}\{\Delta^2 f(a-w)+\Delta^2 f(a)\}+\ldots$$

The above formulae can be written in the form of u_x

after changing $f\ (a+xw)$ to u_x $f(a)$ to u_0 , $f\ (a-w)$ to u_{-1} and so on Proofs of these formulae are given later on

The central difference formulæ are applicable to important problems such as Subtabulation, Estimation of population for individual ages when populations are given in age groups, inverse interpolation, and derivatives of a function The detailed account will be found in Calculus of observation by Whittaker and Robinson, Chap IV

Graduation

Let u_1, u_2 u_n be the set of values as a result of observation or experience, corresponding to equidistant values of the argument If these values have been derived from observations of some natural phenomenon, they will be affected by errors of observation, if they are statistical data they will be affected by irregularities arising from the accidental peculiarities of the data If we form a table of the differences $\Delta\ u_1 = u_2 - u_1$, $\Delta\ u_2 = u_3 - u_2$,— it will generally be found that these differences are not regular, so that the difference table cannot be used for the purposes to which a difference table is nsually put namely *for interpolated values of u,* or *differential coefficient* of u, with respect to its argument

Before the difference table is used, we must perform a process of 'smoothing' that is we must find another sequence

u'_1, u'_2, u'_3 u'_n, whose terms differ as little as possible from the term of the sequence u_1, u_2, —u_n, but having regular differences This smoothing p leading to the formation of u'_1, u'_2 .is called ,the graduation or adjustment of the observations or smoothing of the data For example, mortality is a function of age ana if the mortality rates are tabulated at successive ages on the basis of observed numbers living and dying during ye or period of years, the resulting series will show a definite trend, having however, the fluctuations of sampling. The series must be smoothed before it is used for actuarial purposes and it is the object of graduation to remove such disturbances in a systematic manner, without spoiling the observed facts as far as possible Smoothing can be done by using a freehand curve fitting the data and by using the method of moving averages There was several methods of graduation such as of Wool house and of Spencers. Thei are rather difficult to be given here They are dealt with in Calculus of observation by Whittakar and Robinson (See also Yule and Kendell)

Exercise XI.

1. Given the cubes as follows, find the cubes of 32 3 and 33,1.

Number,	31,	32,	33,	34,	35.
Cubes	29791,	32768,	35937,	39304,	42875.

Ans. 33698 267 and 36264·691.

2. Given

x	2 5,	3,	3·5,	4,	4·5,	5.
y	24·145,	22 043,	20·225,	18·644,	17·262,	16·047.

Find for x=2 75, A 23·05

3

	Marks obtained	*Candidates*
Not more than	45	447
	50	484
	55	505
	60	511
	65	514

Estimate the number of candidates securing not more than 48 Marks

Ans 471

4 The following are the annual premiums required by an Insurance Company to secure Rs 1,000 with profits by making twenty payments in all What would be the premium payable at the age of 26 next birthday ?

Age next birthday	Years	20	25,	30,	35
	Rs	36,	39,	42 12	47 6

Ans Rs 39 12 ans (nearly)

5

x	$f(x)$	x	$f(x)$
25	52	40	84 1
30	67 3	50	92 4

Find the approximate value for $x=35$

Ans 77 5

6 *The pressure of wind in pounds per square feet*, corresponding to the Velocity in miles per hour has been determined by experiment to be approximately as follows —

Velocity	10,	20	30	40
Pressure	1 1,	2,	4 4,	7 9

Estimate the pressure for a velocity of 25 miles per hour

Ans 3 03

7 Death Rates per 100,000 population

	Typhoid	*T B*
1906	31·3	157 1
1909	21 1	139 3
1912	16 5	129·8
1915	12 4	127 7

Estimate the Death Rates for 1910

Ans. 19·19 and 135·25.

8

x	5,	7,	11,	13,	17.
$f(x)$	150,	392,	1452,	2366,	5202.

Find the function by Lagrange's formula when the argument has the values 9 and 6 5 respectively

Ans 810 and 316 875.

9

Ages.	*Proportion occupied per 10,000 of total*
10—15 years,	193 5
15—20	880
20—25	933
25—35	1636
35—45	1201
45—55	830

Determine the number under 30 years

Hint—Take cumulatives at 15, 20, 25 etc and apply Lagrange's formula, with $x=30$. In a frequency distribution it is better to take cumulatives

Ans. 2879

10 In the following table h is the height above sea level and p the barometric pressure Calculate p when $h=5280$.

$h=0$,	2753,	4763,	6942,	10593.
$p=30$,	27,	25,	23,	20

(*M A. Aligarh 1942*) *Ans. 24·5.*

11 In Q 2 find the value for $x=4\frac{1}{8}$ by Gauss and Stirling formula

Ans 18 3 nearly

12 Given Sin 45° = 7071 Sin 50 = 766

Sin 55 = 8192, Sin 60 = 866 find Sin 52

(M A 1942) Ans 788

13 Estimates the population in 1925 of a place having the following record —

Year	*Population in thousands*	*Year*	*Population in thousands*
1891	46	1921	93
1901	66	1931	101
1911	81		

(M A 1942) Ans 96˙837

14 Given the data

x,	0,	1,	2,	5,
$f(x)$	2,	3,	12,	147

form the cubic function of x

(M A 1943) Ans x^3+x^2-x+2

15 The population of a country is given in millions

1911	1921	1931	1941
315	319	353	390

Estimate the population for 1936

(B Com 1945) Ans 372 8125

16 Given Sales in thousand as —

1927	1929	1931	1933	1935
230	390	582	799	1035

Find for 1928

B Com Supp 1945) Ans 305˙513

\ -17. *Given*

Year	1835,	1840,	1845,	1850,	1855,
Cost in Rs. 1,000 (y)	5526,	4577,	?	5395,	5890,
	y_0	y_1	y_2	y_3	y_4

The problem is to find the value for 1845. This can be done by the method of differences with the help of the following difference table

y	First difference Δ	Δ^2	Δ^3	Δ^4	Δ^5
y_0					
	y_1-y_0				
y_1		$y_2-2y_1+y_0$			
	y_2-y_1		$y_3-3y_2+3y_1-y_0$		
y_2		$y_3-2y_2+y_1$		$y_4-4y_3+6y_2-4y_1+y_0$	
	y_3-y_2		$y_4-3y_3+3y_2-y_1$		$y_5-5y_4+10y_3-10y_2+5y_1\ y_0$
y_3		$y_4-2y_3+y_2$		$y_5-4y_4+6y_3-4y_2+y_1$	
	y_4-y_3		$y_5-3y_4+3y_3-y_2$		
y_4		$y_5-2y_4+y_3$			
	y_5-y_4				
y_5					

There are four known quantities, so put the fourth lifference Δ^4 equal to zero to have a relation between y's,

$y_4-4y_3+6y_2-4y_1+y_0=0$ putting the values of $y_0, y_1,$

We get $6y_2=4(4577+5395)-(5526+5890)$

Which gives $y_2=4745$ In practice for such questions ake up to fifth differences the data to get approximate esults

18 Given 10, 20, 30, 40, 50,
48 8, 42, 34 4, 27 6. ?
60, 70, 80, 90
14 3, 9 2, 5·5, 3·3.

Hint.—Put $\Delta_4=0$, then taking $y_7-4y_6+6y_5-4y_4+y_3=0$, esult is $y_5=20\ 65$

19. Years 1916, 1918, 1920,
Manufacture of Cement in India in 1000 tons. 39, 84, ?
1922, 1924, 1926.
151, 264, 388
(M. A 1942). Ans. 95·9.

20 Use Bessel's formula to find $f\ (35)$ given $f\ (x)$ 20, 30, 40 and 50 to be 51203, 43931, 34563 and 348.

(I, C. S. 1936). Ans 39431.

21. A student's union had the following numbers of numbers its roll since 1935, 1935 . —36, 37, 37, 39, 40, 1941, 42, 1944, 95, 817,—, 798, 770, 722, ? 707, 711, 746

Use the method of interpolation to make the best timates you can of the numbers in 1937 and 1941.

(C St 1945). Ans. 8.15 8, 705·8.

22 Obtain by a graphical method the missing figu in the following table of one per cent values of chi squ and check the result by an algebraic method

Degrees of Freedom	1,	2,	3,	4,	5,	6,	7
1% chi sq	6 64,	9 21,	11'34,	13 28	×	16 81,	18 48
Degrees	8		9				
	20 07		21'67				

(*Indian Audit and Accounts 1944*).

23 Given the population of a country in millions as —

1901	1911	1921	1931	1941
282	318	339	352	388

Estimate the population for 1936 by an algebraic 1 and also by means of graph and account for the differenc if any

Ans 36 425 (*Indian Audit and Accounts 1945*)

24 Given sales in Rs (000) as 1937, 38, 39, 40, 41, 42, 1943, 800, 850 —, 790, 720, —, 810.

Determine the best estimates for 1939 and 1942

(*II Com Panjab 1946*) *Ans 840, 685*

CHAPTER XIII

ASSOCIATION OF ATTRIBUTES, CONTINGENCY ψ^2 TEST AND GOODNESS OF FIT

The method of correlation described before is to find ıt the relationship between two series having class tervals and frequencies In this chapter we shall briefly al with association between two attributes having *class equencies*

Definitions —Let A B, C, denote the presence of e several attributes, and *a*, *b*, *c*, the absence of those tributes Thus if B represents the attribute blindness' will represent non-blindness' *i e* sight If A stands r deafness, *a* stands for 'non A' *i e* hearing and so on he class all the members of which possess the attribute , is called *the class A*, while all the members of which ssess the attribute B the class B and so on The ımber of observations assigned to any class is known , the frequency of the class or the class frequency ombination of attributes are denoted by grouping together ıe letters that indicate the attributes concerned AB presents the combination of deafness and blindness, *b*A, on blindness and deafness Combination of capital letters B stands for positive attributes and *ab*, for nega ve attributes AB and *ab* are thus pairs of contraries A' which specifies only one attribute is called a class f the first order, AB specifying two attributes, a class f the second order

All the classes of the same order which equal to ıe total number of attributes, form an Aggregate of frequencies of that order Thus ab Ab, ab, aB ' an aggregate of frequencies of the second order Whe no attributes are specified, the total number of ob vations is denoted by n and is reckoned as a frequency order zero While tabulating, class frequencies should arranged so that frequencies of the same order and frefuenc belonging to the same aggregate are kept together I may be noted that, A will denote the number of A's *i e* objects possessing attribute A, a will denote the number a's *i e*, the objects not possessing attribute A on

AB will denote, the number of AB s *i e*, the objec possessing attributes A and B and so on for others

In a table for the case of 3 attributes, twenty sev frequencies will occur, 1 of order zero $= n$ 6 of the order A B C, 12 of the second and 8 of the th
a b c

In general for k attributes, there are 3^k distinct class fre encies, if n is counted

Any class frequency can always be expressed in of class frequencies of higher order Thus $A+a=n$ $B+b=n$, $AB+Ab=A$, $AB+aB=B$, $ab+aB=a=n-A$ and $AB=ABC+ABc$ Every class frequency can thus expressed in terms of the frequencies of the highest *i e*, of order k The classes specified by k attributes *i e* those of the highest order, are termed the ultimate class frequencies Thus

$$A=AB+Ab=ABC+ABc+AbC+Abc.$$

Every class frequency can be expressed as a sum certain of the ultimate class frequencies If the ass-frequencie- are given, the frequencies of the positives asses, including n can be worked from the relations Aven above The number of ultimate class frequencies 2^{n} and the 3^{n} frequency may all be expressed in terms 2^{n} ultimate class frequencies or of the 2^{n} positive class-quencies

Criterian and Tests of Indepennence of Attributes—there is no relationship of any kind between two attri-es A and B, we expect to find the same proportion of amongst the B's as amongst the not-B's. Two such elated attributes may be termed as independent and the terian of independence for A and B is

$$\frac{AB}{B} = \frac{Ab}{b} \quad (1)$$

This criterian may be put into ferent but more convenient form as

$$\frac{AB}{B} = \frac{AB + Ab}{B + b} = \frac{A}{n} \quad (2)$$

From this we have $\frac{AB}{B} = \frac{A}{n}$, $\frac{AB}{A} = \frac{B}{n}$,

$$\text{or } AB = \frac{A \times B}{n}$$

and $\frac{AB}{n} = \frac{A}{n} \times \frac{B}{n}$, which is frequently applied

From (1) also follows that $Ab = \frac{A \times b.}{n}$

The third form of the test of independence is

$$AB \times ab \quad \frac{A \times B \times a \times b.}{n^{2}} \quad aB \times Ab$$

being equal to e same fraction, therefore, from this follows,

$$AB \times ab = aB \times Ab$$

Association of Attributes —Let the attributes A B be not independent but related in some way or Then if $AB > \frac{A \times B}{n}$ A and B are said to be positiv associated or simply associated If $AB < \frac{A \times B}{n}$ A and are negatively associated or simgly disassoc ated (Exercise XII 6) It may be noted that in Statistics butes are to be associated only when they appear to a large number of cases than they are expected to they are independent

To measure the degree of association there several co efficients that have been devised but simplest is

$$Q = \frac{AB \times ab - Ab \times aB}{AB \times ab + Ab \times aB}$$

If $Q=0$ the attributes are independent if $Q=1$ t are completely associated and if -1 they are com disassociated The attributes can be put in the form Table with either A or B on the column or row having two rows and two columns thus (2×2 fold)

	Attribute		
Attribute	B	*b*	*Total*
A	AB	A*b*	A
a	*a*B	*ab*	*a*
Total	B	*b*	*n*

Contingency Tables and ·Co-efficient of Contingency.—Let the classification of the attribute A be s-fold nd that of B's *t* fold There will be *st* classes of the ype $A_l B_m$ (l and m may take any values 1, 2, . .). et the frequencies of A's be denoted by (A_l) and of B's y (B_m) and of $A_l B_m$ by $(A_l B_m)$ and so on The data an be set out in the form of a table of *t* rows and columns. The table described above is fourfold (2×2 lassification)

A general contingency table is of the form ($s \times t$ fold)

Attributes			A			Total
		A_1	A_2		A_s	
	B_1	(A_1B_1)			$(A_s B_1)$	(B_1)
B	B_2	(A_1B_2)				
						.
	B_t	$(A_1B_t$			(A_sB_t)	(B_t)
	Total	(A_1)			(A_s)	n

The frequency of any class $A_l B_m$ is entered in the Compartment or cell common to the l-th column and the m th row The frequency falling in a call is said to be the cell frequency. If A and B are completely independent for all values, the $(A_l B_m) = \frac{(A_l\ )\times B_m)}{n} = (A_l B_m)'$.

If A and B are not completely independent $(A_l \; B_m)$ and $(A_l \; B_m)'$ will not be indentical Let the $(A_l \; B_m) - (A_l \; B_m)'$ be denoted by d_{lm}.

The co efficient of association or the co efficient of square contingency is given according to Pearson by,

$$C = \sqrt{\frac{\psi^2}{n+\psi^2}} \quad \text{where } \psi^2 = \Sigma \left(\frac{d^2_{lm}}{(A_l \quad B_m)'} \right)$$

ψ^2 (chi square) is square contingency and $(A_l B_m)'$ $= \frac{A_l \times B_m}{n}$.

A simpler form due to Yule is $C = \sqrt{\frac{S-n}{S}}$ where $S = \Sigma \frac{(A_l \quad B_m)^2}{(A_l B_m)'}$, where A_l B_m are the actual obesrved frequeneies.

It is desirable for the calculation of C, to use a (5×5 fold classification.

General procedure —To find ψ^2, first of all cal the frequency which would be expected in each cell on null hypothesis i.e., on the assumption that the two attr butes are not associated with another at all i.e., $(A_l \; B_m)'$ for all l and m. Subtract this ex frequency from the actual observed frequency in each cell square these differences and divide by the frequency $(A_l \; B_m)'$ to get ψ^2. If the null hyqothesis correct ψ^2 and C will be zero To test the contingency the Calculated Value of ψ^2 may be compared with the Values given by Fisher (Fisher and Yates Table, IV) for N degrees of freedom for Probability P = 05 (5 percent level of significance) If the Calculated Value is greater, then the Nul hypothesis is disproved and thus there is a significant association The degree of freedom are given by $N = (s-1)(t-1)$ where s is the number of columns and t the number of rows (e g See Exercise XII 9)

Table of ψ^2 (5 per cent level of significance) P = 05

N	ψ^2	N	ψ^2	N	ψ^2
1	3 841	11	19 675	21	32 671
2	5 991	12	21 026	22	33 924
3	7 815	13	22 362	23	35 172
4	9 488	14	23 685	24	36 415
5	11 070	15	24 996	25	37 652
6	12'592	16	26 296	26	38 885
7	14 067	17	27 587	27	40 113
8	15 507	18	28 869	28	41 337
9	16 919	19	30'144	29	42 557
10	18 307	20	31 41	30	43 773

Goodness of Fit —The ψ^2 distribution leads to tests of the correspondence between theory and fact and is described as a test of the goodness of fit If an observed frequency distribution of a variable is given and we want to examine the validity of some hypothesis about it, this can be done by calculating the expected or theoretical frequencies and examining the agreement or goodness of fit' of the observed and theoretical frequencies with the help of $\psi^2=\Sigma \frac{(f'-f)^2}{f}$ where f' denotes the observed or actual frequencies and f, the theoretical frequencies The *whole working is the same* as for ψ^2 in contingency described above The value of ψ^2 may be compared from the Tables (Fisher and Yates) Further if the probability is very low, it will mean a poor fit, if high then the fit is excellent and so on (See Exercises No. 13).

Contingency tables with Small Frequencies,
Yates corrections

If the number of frequencies in one or more compartments of the table is small (less than 5) certain changes have to be made to obtain better results

Yates correction is made in the smallest frequency, i e, add $\frac{1}{2}$ to the smallest frequency in the contingency table and adjust other frequencies so that the marginal totals *remain the same.*

e g,

	Attacked by disease	*Not attacked*	
Inoculated	10	3	*13*
Not Inoculated	2	5	7
	12	8	20

In the table the frequencies according to Yates correction are changed to

9·5	3·5	13
2·5	4·5	7
12	8	20

The rest of the procedure is the same as for ψ^2 distribution In the above example $\psi^2=2·6965$. Comparing from he tables for one degree of freedom the result is not significant.

Exercise XII.

1 If A and B are independent attributes, how many AB will there be in 1000 observations if there are 100 A's and 400 B's ? What will be the number of ab's ?

Sol —Using $AB = \frac{A \times B}{n} = \frac{100 \times 400}{1000} = 40.$

Again $ab = \frac{a \times b}{n} = \frac{900 \times 600}{1000} = 540.$

2 Given the actual observations as, A (Vaccinated people) $= 30$, B (*not attacked by small-pox*) $= 60$, $n = 150$ AB (people who were Vaccinated and not attacked by small-pox) $= 12$ Are the attributes A (Vaccination) and B (exemption from attack) independent ?

Ans. Yes, *i e.* Vaccination and exemption are not related at all.

3. In Q. 2, given ab (people not vaccinated and attacked) $= 58$, are a and b independent ? *No.*

4. In Q. 2 if $AB = 15$; $ab = 68$, $Ab = 20$, $aB = 51$, are A and B independent ? Use the test, $AB \times ab = aB \times Ab$ *Yes.*

5 If the second order frequencies have the values, $AB = 110$, $aB = 90$, $Ab = 290$, $ab = 510$ test the independence of A and B *Ans No.*

6 The attributes in Q 2 are placed in the form ᵕ table as

	A	a	n
B	f0	10	70
b	20	10	30
n	80	20	100

Test the association

Sol—Applying the test of association $AB > \frac{A \times B}{n}$

we have 60 is greater than $\frac{80 \times 70}{100}$ hence the vaccination exemption from attack are positively associated

Also $ab = 10$ and is greater than $\frac{a \times b}{n}$ i e $\frac{20 \times 30}{100}$, and so a and b also possess positive asso˙˙ For Attributes A and b we find $Ab = 20$ is less $\frac{A \times b}{n}$ i e $\frac{80 \times 30}{100}$ they are negatively

Similarly a and B are disassociated

7 Test association between injection against typhoı and exemption from attack from the contingency table

	A *Not attacked*	a *Attacked*	
B Injected	270	10	280
b Not Inj	480	60	540
	750	70	820

Ans *Associated positıı*

8 Determine the Co efficient of association for Q 6

Ans

9 Find ψ^2 and test for association the following data

	A_1	A_2	A_3	
B_1	215	325	60	600
B_2	135	175	90	400
	350	500	150	1000

Sol —First construct the table for expected frequen-
:s *i e* , of Independence Values by finding

$$(A_l \ \ B_m)' = \frac{A_l \times B_m}{n}, \text{ table is}$$

A_1	A_2	A_3	
210	300	90	600
140	280	60	400

thus $140 = \frac{350 \times 400}{1000}$ and

$$300 = \frac{500 \times 600}{1000}$$

$$\psi^2 = \frac{(215-210)^2}{210} + \frac{25^2}{300} + \qquad \frac{(90-60)^2}{60} = 30\ 5$$

arly Degrees of freedom are $(3-1)(2-1)=2$ and for s from the table $\psi^2 = 5\ 9$ The calculated value is much eater than this value hence the Null hypothesis departs nificantly from independence and there is significant sociation and $C = \sqrt{\frac{(30\ 5)}{1000+(30\ 5)}}$

10 Is there a significant association between A and B om the following (2×2) table ?

	A_1	A_2	
B_1	*a*, 64	*b*, 26	90
B_2	*c*, 21	*d*, 49	70
	85	75	160

Sol — For (2×2) table ψ^2 can also be determined fron the formula

$$\psi^2=\frac{(a+b+c+d)(ad-bc)^2}{(a+b)(a+c)(b+d)(c+d)} \quad \textit{Ans Yes}$$

11 Determine the co efficients Q and C for the ta(and compare these with Yule's co efficient C

Attributes	1	2
1	450	242
2	36	272

Ans.' Q= 86 and C=

12 Given the following contingency table for Ha(Colour (5 categories) and Eye Colour (5 categories). F the valu^ of C. Is there good association ?

0	0	2	10	11
1	13	69	189	13
5	96	336	91	6
22	89	32	5	0
3	6	1	0	0

Ans 73, yes

13. Examine the goodness of fit for the follow(frequency distributions, total 398 and degrees of freedom 8

	0	1	2	3	4
Actual frequencies	15,	38,	76,	70,	64,
Theoretical	8 2,	32,	61'5,	80,	77'7,
	5	6	7	8	9
Actual	53,	31,	19,	14,	20
	60'5,	39'2,	22,	10'6,	7'3

Sol —ψ^2 is greater than ψ^2 in the tab'e for 8 degrees o freedom, and so probability is very sm [illegible] the [illegible] eries is a bad fit to the observed distribution, Tables exist iving ψ^2 for different values of probability P. Generally : P is less than the selected fiducial limit of ·05 or of ·01, he hypothesis is said to be false

14. Given the following Actual and theoretical Normal requencies (total 400) Test the goodness of fit (degress f freedom 10)

ctual	4	11	17	29	43	56
heoretical	4·6,	7·0,	16·8,	30 3,	44·7,	59 1,
	58	63	61	25	20	9
	65,	60 4,	47 5,	31 16,	18 25,	8 8
	4					
	5·3.					

Ans Good

15. The table given below shows the data obtained uring an epidemic of cholera

	Attacked	*Not Attacked*	*Total*
Inoculated	31	469	500
Not Inoculaied	··· 185	1315	1500
	216	1784	2000

Test the effectiveness of inoculation in preventing the ttack of cholera.

[Five per cent value of ψ^2 for one degree of freedom 3·84]

(Indian Audit and Accounts Service 1941)

Ans Significant

16 Discuss the resemblance of stature of parent and offspring from the following —

Parent

Offspring	*Very Tall*	*Tall*	*Medium*	*Short*	*Total*
Very Tall	20	30	20	2	72
Tall	14	125	85	12	236
Medium	3	140	165	125	433
Short	3	37	68	151	259
Total	40	332	338	290	1000

(*I C S 1936*) *Ans Great.*

17 The following table shows the association, among 1000 school boys, between their general ability and their mathematical ability Calculate the coefficient of contingency between the two

General ability

Math ability	*Good*	*Fair*	*Poor*
Good	44	22	4
Fair	265	257	178
Poor	41	91	98

M A (*Maths 1945*)

Ans $\psi^2 = 68\ 8$ $C = 24$

18 In an experiment on the immunization of [illegible] from anthrax the following results were obtained Der your inference on the efficacy of the vaccine

	Died of anthrax	*Survived*
Inoculated with vaccine	2	10
Not inoculated	6	6

(*Indian Audit & Accts 1943*

19. The following table gives the results of a series of controlled experiments Discuss whether the treatment may be considered to have any positive effect

		Positive	No effect	Negative
Treatment	–	9	2	1
Control	–	3	6	3
Total	–	2	8	4 = 24

(Indian Audit 1944)

20 The table below shows the data obtained during an epidemic of cholera

	Attacked	*Not attacked*
Inoculated	30	470
Not inoculated	185	315

Test the effectiveness of incculation in preventing the attack of cholera

(C. St 1945)

21 Explain the use of the Tests of significance and of association in the analyses of commercial data

(M Com 1946)

CHAPTER XIV

CORRELATION RATIO PARTIAL AND MULTIPLE CORRELATION

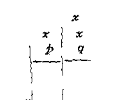
The methods ot measuring correlation described bef are useful when the regressions of the two variables t each other are linear If regression is non linear the degre of association is measured by means of the Correlati Ratio

There are two Correlation rat os for each pair of v r ables x and y explained below

Let n_p = the number of y s in any array x_p

y_p is any y in array x_p

$\overline{y}_p$ the mean of y s in any array x_p



y the mean of all the y s σ^2_p the variance of y s array x_p

σ^2_y the variance of all the y s

then the Correlation Ratio $\eta^2_{yx} = \Sigma \frac{\left\{n_p\left(y_p - y\right)^2\right\}}{n\ \sigma^2_y}$

where n is the total number of frequencies for the whole listribution.

$$\text{Similarly} \quad \eta^2 x_y = \Sigma \frac{\left\{ n_{p'} \left(x_{p'} - x \right)^2 \right\}}{n \sigma^2_x}$$

For application of it see Exercise No 1

Correlation Ratio is, in fact, the ratio between the tandard deviation of the means of arrays and the standard leviation of the whole sample and is chiefly used when the lata are numerous and can be arrayed in the form of a *Correlation Table*

In finding η_{y_x}, the numerical values of x variates ire not used, hence it is possible to find correlation ratio when only one set of variates is quantitative, the others nay be attributes such as eye colour intellectual qualities. The co efficient of correlation r, cannot be found when one variate is qualitative, for that we must have both quantitative *High correlation is associated with values of η approaching unity*

When the frequency distribution is normal the correlation ratio is identical with the correlation co efficient r

The Standard Error of η is $\dfrac{1-\eta^2}{\sqrt{n}}$

Partial Correlation—Sometimes it may be desired o measure the relationship between the independent and he dependent with the effect of other independent variables held constant or eliminated Two variables x and y are correlated partly on account of the fact that each of them is correlated with a third variable z We may be required to find the correlation between x and y, quite apart from the influence of z This is done by the method of partial correlation for instance, yield of a crop of cereals depends partly on rainfall partly on sunshine and other conditions The relationship between yield and rainfall can be worked keeping other condition constant

The correlation between x and y with the effect of z unchanged or ignored, is given by the co efficient of partial correlation (of first order)

$$r_{xy\,z} = \frac{r_{xy} - r_{xz}\,r_{yz}}{\sqrt{1-r^2_{xz}}\;\sqrt{1-r^2_{yz}}}$$

Or briefly $r_{12\,3} = \dfrac{r_{12} - r_{13}\,r_{23}}{\sqrt{1-r^2_{13}}\;\sqrt{1-r^2_{23}}}$.

Where $r_{xy\,z}$ means the correlation between x and y, z being constant, r_{xy} is the correlation between x and y, r_{xz} between x and z r_{yz} between y and z

If there are four variables, the co efficient of 2nd order is (keeping z and u constant)

$$r_{xy\,zu} = \frac{r_{xy\,z} - r_{xu\,z}\,r_{yu\,z}}{\sqrt{1-r^2_{xu\,z}}\;\sqrt{1-r^2_{yu\,z}}}$$

Or briefly $r_{12\,34} = \dfrac{r_{12\,3} - r_{14\,3}\,r_{24\,3}}{\sqrt{1-r^2_{14\,3}}\;\sqrt{1-r^2_{24\,3}}}$

In this way we can proceed to several variables

Example —Three tests in mathematics were given to a group of students and three sets of scores were correlated with each other, giving $r_{xy} = {}^{\cdot}6, r_{xz} = 5,\ r_{yz} = {}^{\cdot}4$.

What is the correlation between first and second keeping the third constant ?

$$r_{xy\,z} = \frac{6 - 5 \times {}^{\cdot}4}{\sqrt{1-25}\ \sqrt{1-16}} = \frac{4}{{}^{\cdot}79} = 5.$$

Multiple Correlation —In simple correlation we dealt with the relationship between the derendent variable and a simple independent variable

Multiple Correlation is a measure of the combined effect of two or more independent variables upon one dependent variable For instance, the simple co efficient of correlation between rainfall during a certain period and yield of corn is less than 1. This clearly indicates that some other factor and factors must be taken into account if we want to measure the effect of all the independents upon the yeild of corn. The co efficient of multiple correlation is a numerical expression of the extent to which one dependent variable is related to or influenced by the joint or total effect of two or more factors Multiple Correlation is also known as multivariate correlation, just as simple correlation is called bivariate correlation

Let $\sigma_1, \sigma_2, \sigma_3, \sigma_4$ denote the standard deviations of the variables or characters 1, 2, 3, 4, $(x_1, x_2, x_3, x_4$, say) then σ_{12} is the standard deviation of first one character x_1 when the influence of second x_2 is kept constant, $\sigma_{1\,23}$ the

S deviation of fisrt when the influence of 2nd and 3rd x_2 ard x_3 is kept contant

$$\sigma_{1\ 2}=\sigma_1\sqrt{1-r^2_{12}}\quad a_{1\ 23}=\sigma_1\sqrt{1-r^2_{13}}\sqrt{1-r^2_{2\ 3}}$$

$$=\sigma_1\sqrt{1-r^2_{12}}\sqrt{1-r^2_{13\ 2}}$$

$$\sigma_{1\ 234}=\sigma_1\sqrt{1-r^2_{1L}}\sqrt{1-r^2_{13\ 2}}\quad\sqrt{1-r^2_{14\ 23}}$$

$$\sigma_{1\ 234}\quad -n=\sigma_1\sqrt{1-r^3_{12}}\sqrt{1-r^2_{13\ 2}}\quad\sqrt{1-r^2_{14\ 23}}$$

$$\sqrt{1-r_{1n\ \ 234}}\quad(n-1)$$

Similarly other standard deviations can be written by analogy such as

$$\sigma_{2\ 31}=\sigma_2\sqrt{1-r^2_{23}}\sqrt{1-r^2_{21\ 3}}$$

The standard error of an estimate of x from a regression equation is $\sigma_{1\ 234}\quad n$

The co efficient of mu tiple correlation is given by

$$R^2_{1\ 23}\quad n=1-\frac{\sigma^2_{1\ 234}\quad n}{\sigma_1^2}$$

Where $R_{1\ 23}$ is the co efficient of the chararcter (1) with the character 2 3 4 n

Exercise XIII

I —Find the correlation Ratio for the following table (Dawson)

	11	10	9	8	7	6	5	4	3	2	1
6	4	1									
5	18	6	1								
4	27	14	5	1							
3	26	15	6	1	1						
2	12	26	7	3	1	1					
1	5	29	13	4	3	1	0	1	0	1	
0	4	13	22	17	8	2	2	2	0	1	1
−1	2	5	17	19	14	5	4	3	3	2	3
−2	·1	2	7	15	20	10	10	7	5	5	6
−3	0	1	3	3	13	20	18	10	12	9	12
−4			2	1	6	12	12	13	16	15	11
−5				1	2	5	4	6	8	5	4
−6						1		1	1	2	1

Sol.—The Values for n_p for each column in x are 99, 112, 83 40,38. The Means *i e*, $\bar{y}_p$ corresponding to each array of x are 3 28, 1·839, ·265, −·769, −1 75, −2·824, −2 92 −3 116, −3 533, −3·375, −3 184 and $\bar{y} = -\ 674$ and $\sigma_y = 2\ 87$

Find $y - \bar{y}_p$, square and multiply by the corresponding Values, n_p, the total sum is 4081·1.

$$\therefore \eta_{yx} = \sqrt{\frac{4081\ 1}{700 \times (2\ 87)^2}} = 84.$$

II.—	5	4	3	2	1	0	1	2	3	4
200−210	...				1		1	6	5	1
190−200				2	4	8	9	9	4	4
180−190				3	8	15	24	9	3	
170−180			2	9	14	24	17	7		
160−170			5	11	19	17	9	3		
150−160		1	4	11	12	13	5	2		
140−150		2	2	9	8	4	2			
130−140			2	3	2	..	1			
120−130	1		1	1						

Total of frequencies = 339.

Ans ·6 *nearly.*

III —Given the following: r between supply and price of a commodity = − 9, r between exports and price = −·75, r between supply and exports = ·6. Calculate the net relationship between supply and price with the effect of exports held constant.

Ans −·8

IV.—Given the following correlations for intelligenc tests, of 200 students, $r_{ia} = 41$, $r_{is} = 71$ $r_{as} = 5$ where i denotes intelligence test score, a, age, s, schola achievement.

Find $r_{is.a}$, $r_{ia.s}$, $r_{as\,i}$ and their Probable Errors *Ans.* '*63*, *09*, *3* 3

Hint.—Probable errors are sound in the same way *for simple r*

V—Given, $r_{11\,3} = -\ 34$, $r_{14\,3} = -\,'43$ $r_{24\,3} = '187$; $r_{14\,2} = -\ 43$ $r_{13\,2} = -\,'62$, $r_{34.2} = 2$ find the values *of* $r_{12\,134}$ – $r_{14\,23}$ (approx) (answers in all the questions are approxim in decimals) *Ans.* $-$'*29 and* $-$.4

VI.—The Correlation between the intelligence-ratios and height of 406 students was 24, that between their and height was 85 and the correlation between the age intelligence-ratio was '007. Taking height as 1, age 3, intelligence-ratio 2 $\sigma_1, = 15\ 22$, find the standard error of estimate and co-efficient of multiple Correlation

Sol.—First of all find the partial no efficients.

Which are $r_{12\,3} = 46$, $r_{23\,1} = -\ 415$, $r_{13\,2} = '88$, then $\sigma_{1\,23} = 6'9$, $R^2 = 97$ and $R = 98$

VII.—Given $\sigma_1 = 44'7$; $r_{12} = -\ 5$ $r_{13\,2} = -.62$, $r_{14\,23} = -\ 405$ find $\sigma_{1\,234}$ and R *Ans.* *4 13 and 78.*

VIII.=For three 'Variables given $\sigma_1 = 1\ 7$, $\sigma_2 = 1\ 3$ $\sigma_3 = 3\ 1$, $r_{12} = -\ 65$, $r_{13} = -\ 13$, $r_{23} = 6$ Calculate the C efficient $R_{2\,23}$, $R_{1\,23}$ and $R_{3\,12}$.

Ans. '*75* '*85* *a* *4* '*7*

IX —Given for three agricultural products r_{12}='24, $_{13}$= 85 $r_{\ 3}$= 07 σ_1=15 2 determine $R_{1\ 13}$ (C st 1945) Hint see Ex VI

X =Find the Regression equations for Q VIII.

Sol —The regression equations will be for 3 variables, x_1, x_2 and x_3 $x_1 = b_{12\ 3}\, x_2 + b_{13\ 2}\, x_3$

where $b_{12\ 3} = r_{12\ 3}\ \dfrac{\sigma_{1\ 23}}{\sigma_{2\ 13}}$ etc

$x_2 = b_{23\ 1}\, x_3 + b_{21\ 3}\, x_1 \quad x = b_{31\ 2}\, x_1 + b_{3\ \ 1}\, x_2$

Ans $x_1 = -1\ 2\ \ x_2 + 23x_3 \qquad x_2 = -\ 44\ \ x_1 + 2\ \ x_3$, $x_3 = 84\ x_1 + 2\ 2\ x_2$

CHAPTER XV

MATHEMATICAL THEORY OF INTERPOLATION

We have already exolained Interpolation and its me.hod of calculation in Chapter XII

In this chapter mathematical proofs will be given

Symbolic operators The Interpolation formulæ can be expressed in terms of operators Δ, and E, defined as follows for a function $f(a+xw)$ with equal intervals w

$\Delta\ f(a) = f(a+w) - f(a)$, known as the first difference of $f(a)$

$$\begin{aligned}\Delta^2 f(a+w) &= \Delta\ f(a+w) - \Delta\ f(a)\\ &= f(a+2w) - f(a+w)\\ &\quad -\{\ f(a+w) - f(a)\}\\ &= f(a+2w) - 2f(a+w) + f(a)\end{aligned}$$

known as the second difference

$\Delta^3 f(a+w) = \Delta^2 f(a+w) - \Delta^2 f(a) - \sigma^2 f(a)$ and so on

Diff'rence Table —The differences are arranged in the

$$= f(a+nw) - nf(a+nw-w) +$$
$$\frac{n(n-1)}{2!}\,f(a+uw-2w) - \frac{n(n-1)(n-2)}{3!}\,f(a+nw-3w)$$
$$+ \quad . \quad + (-1)^n f(a)$$

Again since

$f(a+xw) = E^x f(a) = (1+\Delta)^x f\,a)$ we have after expanding as binomial

$$f(a+xw) = f(a) + x\Delta^3 f(a) + \frac{x(x-1)}{2}\Delta x f(a)$$
$$+ \frac{x(x-1)(x-2)}{3!}\Delta^3 f(a) + \quad - \quad \Delta^x f(a)$$

which expresses the function terms of $f(a)$ and the successive differences of $f(a)$

It is customary to denote $n\,C_r$ as (n_r)

Example To express $\Delta^3 f(a)$ in terms of the functions and $f(a+3w)$ in terms of successive differences. $\Delta^3 f(w)$ $= f(a+3w) - 3f(a+2w) + 3f(a+w) - f(a)$

$f(a+3w)$ in terms of $f(a)$ and successive differences can be written as $f(a+3w) = f(a) + 3\Delta f(a) + 3\Delta^2 f(a) + \Delta^3 f(a)$ which can be verified directly by the definition of Δ, Δ^2, and Δ^3

If a function is in the form $u_{x,\, w=1}$ then $\Delta u_x = u_{x+1} - x^u$ e g, $\Delta x^3 = (x+1)^3 - x^3$.

$$\Delta^2 u_x = \Delta u_{x+1} - \Delta u_x \text{ and so on}$$

Differences of a Polynomial and factorial Polynomials

Theorem If there is a polynomial of nth degree, prove that nth differences are constant and $(n+1)$th differences zero

Proof Consider the polynomial

$$f(x) = Ax^n + Bx^{n-1} + Cx^{n-2} + \quad . \quad . + K$$
$$\Delta f(a) = ,(a+w) - f(a)$$
$$= A\{(a+u^n - a^n\} + B\{(a+w)^{n-1} - a^{n-1}\} + .. \quad (1)$$

Using the Binomial theorem,

$(a+w)^n = x^n + nwx^{n-1} + (n_2)w^2a^{n-2} +$

From (1)

$\Delta f(a) = A\{nwa^{n-1} + n_2 w^2 a^{n-2} + \quad w^n\} + B\,(n-1)wa^{n-2} + \quad \} +$ which is a polynomial of degree $(n-1)$ in a. The first differences of a polynomial thus represent another polynomial of degree less by one. Proceeding in this way we derive that nth differences represent a polynomial of degree $(n-n)$ or are constant and the $(n+1)$th differences are zero.

Tho polynomial $x(x-1)(x-2)\ldots(x-r+1)$ is denoted by $\overset{r}{[x]}$ or $\underset{r}{x}$ or $x^{(r)}$ and may be called factorial polynomial.

Thus $[a]^r = a(a-1)(a-2)\ldots(a-r+2)(a-r+1)$.

$[a+1]^r = (a+1)a(a-1)(a-2)\ldots(a-r+3)(a-r+2)$

$$\begin{aligned}\Delta[a]^r &= [a+1]^r - [a]^r \\ &= a(a-1)(a-2)\ldots(a-r+2)\{a+1-(a-r+1)\} \\ &= r[a]^{r-1}\end{aligned}$$

Or in general $\Delta[x]^n = n[x]^{n-1}$ which corresponds to $\frac{d}{dx}x^n = nx^{n-1}$ in differential calculus.

Theorem Show that every polynomial can be expressed in terms of factorial polynomials.

Proof Let $f_n(x)$ be a polynomial of nth degree. Dividing by x, $f_n(x) = a_0 + [x]f_{n-1}(x)$ where $f_{n-1}(x)$ is a polynomial of degree $(n-1)$.

Dividing further by $x-1$, $x-2$, $x-3$ ... We shall obtain the required result

$\underset{n}{f}(x) = a_0 + a_1[x]f_{n-1}(x) + a_2[x]^2 f_{n-2}(x) + \ldots$

Example Express $y = 2x^3 - x^2 + 3x - 1$ in factorial notation

Dividing by x

$y=(2x^2-x+3)[x]+1$

Dividing further by $x-1$ $\quad y=1+(2x+1)x(x-1)+4x$

$$=1+(2x+1)[x]^2+4[x]$$

Dividing by $(x-2)$ $y=1+4[x]+5[x]^2+2[x]^3$

$\Delta f(x)=4+10[x]+6[x]^2$

$\Delta^2 f(x)=10+12x$

$\Delta^3 f(x)=12$

$\Delta^4 f(x)=0$

Proof of Newton s formula for equal intervals Let be one of the tabulated values of the argument of a polynomial of degree n

The polynomial $f(a+xw)$ can be written as $f(a+$
$=a_0+a_1[x]+a_2[x]^2+a_3[x]^3+ \qquad a_n[x]^n$

Differencing,

$\Delta f(a+xw)=a_1+2a_2[x]+3a_3[x]^2+4a_4[x]^3+\dots$

$\Delta^2 f(a+xw)=2a_2+2\ 3\ a_3[x]+4\ 3a_3[x]^2+$

$\Delta^3 f(a+xw)=2\ 3a+4\ 3\ 2a_3[x]+$

The value of the co efficients a_0 a_1, a_2

will be determined by putting $x=0$ in the above

Thus $a_0=f(a)$

$$a_1=\Delta f(a)$$

$$a_2=\frac{\Delta^2 f(a)}{2!}$$

$$a_3=\frac{\Delta^3 f(a)}{3!}$$

$$a_4=\frac{\Delta^4 f(a)}{\underline{|4}}$$

$$a_n=\frac{\Delta^n f(a)}{\underline{|n}}$$

$$f(a+xw)=f(a)+x\Delta f(a)+\frac{x(x-1)}{2!}\Delta^2 f(a)+\frac{x(x-1)(x-2)}{3!}$$
$\Delta^3 f(a)+ \quad - \quad - -$

This is known as Gregory Newton formula or simply ewton formula of interpolation which is used for interpotin and can be represented geometrically as a straight ne or a curve

For examples in New on s formula see Chapter XII and xercise XI

Divided Differences and Newton s formula for unequal itervals

Let a, b, c, d be arguments at unequal intervals nd $f(a)$ $f(b), f(c)$ $- -$ the corresponding functions

The divided difference of the first order is defined by

$$ab)=[ab]=\frac{f(a)}{a-b}+\frac{f(b)}{b-a}=\frac{f(a)-f(b)}{a-b}$$

The divided difference of second order is defined by

$$abc)=\frac{f(a)}{(a-b)(a-c)}+\frac{f(b)}{(b-a)(b-c)}+\frac{f(c)}{(c-a)(c-b)}$$
$$=\frac{f(ab)-f(bc)}{a-c}$$

$$f(abcd)=\frac{f(a)}{(a-b)(a-c)(a-d)}+\frac{f(b)}{(b-a)(b-c)(b-d)}$$
$$+\frac{f(c)}{(c-a)(c-b)(c-d)}+\frac{f(d)}{(d-a)(d-b)(d-c)}=\frac{f(abc)-f(bcd)}{a-d}$$

Similarly the divided differences of higher order can be efined

Newton s formula for unequal intervals

Let $f(x)$ be a function whose divided differences of order ur are zero or so small as to be neglected, such that the ird order differences are constant The table of divided

$$\begin{array}{l|lllll}
a & f(a) & & & & \\
b & f(b) & f(\,rb) & f(abc) & f(abcd) & f(abcde) \\
c & f(c) & ,(bc) & f(bcd) & f(bcde) & \\
d & f(d) & f(cd) & f(cde) & & \\
e & f(e) & f(de) & & &
\end{array}$$

The problem is to find the value of the function for other argument u, which may or may not be contained in the above arguments,

Since the divided differences of third order are constant

$$\therefore \quad f(abcd)=f(uabc)$$

It is known that

$$f(u\,abc)=\frac{f(u\,ab)-f(abc)}{u-c}=f(abcd)$$

$$\therefore \; f(u\,ab)=f(abc)+(u-c)f(abcd)$$

$$\text{Also } f\,u\,ab)=\{f(uc)-f(ab)\}\;\frac{1}{u-b}.$$

$$\therefore \; f\,ua)=f(ab)+u-b)\{f(abc)+(u-c)f(abcd)\}$$

$$\text{But } f(ua)=\frac{f(u)-f(a)}{u-a}$$

$$\begin{aligned}
\therefore \; f(u) &= (u-a)f(ua)+f(a) \\
&= f(a+(u-a)f(ab) \\
&\quad +(u-a)(u-b)f(abc)+(u-a)(u-b)(u-c)f(abcd
\end{aligned}$$

The formula holds in general when the differences of $(n+1)$ the order are zero or negligible and is known as Newton's formula for unequal intervals

Lagrange s formula of interpolation

Let $f(x)$ be polynomial of degree n which for val $a_0, a_1, a_2 \quad \ldots \quad a_n$ of x possess the valuss $f(a_0)$, $f(a$ respectively

The divided differences of order n of a polynomial constant, since the divided differences of order n of each the terms whose degree is less than n is zero The di differences of order $(n+1)$ will be zero.

The divided differences of $(n+1)$ th order, by definition s given by $f(x, a_0 a_1 a_2 \quad a_n)$

$$= \frac{f(x)}{(x-a_0)(x-a_1) \quad -a_n)}$$

$$+ \frac{f(a_0)}{(a_0-x)(a_0-a_1) \quad (a_0- \quad)}$$

$$+ \frac{f(a_1)}{(a_1-x)\, a_1-a_0) \quad (a_1-a_n} +$$

$$+ \frac{f(a_n)}{(a_n-x)(a_n-a_0) \cdot \; \cdot (a_n-a_{n-1})} = 0$$

Which gives on multiplication throughout by $-a_0)(x-a_1) \quad (x-a_n)$

and simplifying

$$f(x) = \frac{(x-a_1)(x-a_2) \cdot \quad (x-a_n)}{(a_0-a_1)(a_0-a_2)\ldots \quad (a_0-a_n)} \; f(a_0)$$

$$+ \frac{(x-a_0)(x-a_2) \quad (x-a_n)}{(a_1-a_0)(a_1-a_2)\ldots \quad (a_1-a_n)} \; f(a_1)$$

$$+ \frac{(x-a_0)(x-a_1)(x-a_3) \quad (x-a_n)}{(a_2-a_0)(a_2-a_1) \quad (a_2-a_n)} f(a_2) + \quad \ldots$$

Which is Lagrange s formula of interpolation of for unual intervals

I or illustration see chapter XII article 2.

Central difference formulae of Interpolation

In the central difference formula, the argument a is taken the centre or near about the centre of the arguments as

Arguments	*Functions*
$a-2w$	$u_{-} = f(a-2w)$
$a-w$	$u_{-1} = f(a-w)$
a	$u_0 = f(a)$
$a+w$	$u_1 = f(a+w)$
$a+2w$	u_2

A Notation δ is used according to which $\delta = \Delta E^{-\frac{1}{2}}$ or $= \delta E^{\frac{1}{2}}$ Thus $\Delta u_0 = \delta u_{\frac{1}{2}}$, $\Delta^2 u_0 = \delta^2 u_1$

The following are the well known formula in Central differences, which will be proved

Newton Gauss formula, known as Gauss formula

Newton Stirlings formula Bessel's and Everett's formulae Gauss formula

Let the function $f(a+xw)$ have the arguments $a-3w$, $a-2w$, $a-w$, $a+w$, $a+2w$ In Newton's formula for unequal intervals, write

$u=a+xw$, $b=a+w$, $c=a-w$, $d=a+2w$, $e=a-2w$ and so in

$$\therefore\ f(a+xw)=f(a)+xw\,f(a,a+w)+xw\,(a+xw-a-w)$$
$f(a, a+w, a-w)+$

But $f(a,a+w)=\dfrac{f(a+w)-f(a)}{w}=\dfrac{1}{w}\Delta f(a)$

$$f(a,a+w,a-w)=\frac{\Delta^2}{2!\,w^2}f(a-w)$$

$f(a,a+w,a-u,a+2w)=\dfrac{1}{3!\,w^3}\Delta^3 f(a-u)$ and so on

Hence

$$f(a+xw)=f(a)+x\Delta f(a)+\frac{x(x-1)}{2!}\Delta^2 f(a-w)$$
$$+\frac{(x+1)x(x-1)}{3!}\Delta^3 f(a-w)$$
$$+\frac{(x+1)x(x-1)(v-2)}{4!}\Delta^4 f(a-2w)$$
$$+\frac{(x+1)x(x-1)(x-2)}{5!}\Delta^5 f(a-2w)+$$ which is Gauss formula

Newton Stirling's formula

In Gauss formula, the terms may be arranged as

$$f(a+xw)=f(a)+x\,[\Delta f(a)-\tfrac{1}{2}\Delta^2 f(a-w)]$$

$$+\frac{x^2}{2!}\Delta^2 f(a-w)+\frac{x(x^2-1^2)}{3}\left[\Delta^3 f(a-w)-\tfrac{1}{2}\Delta^4 f(a-2w)\right]$$

$$+\frac{x^2(x^2-1^2)}{4!}\Delta^4 f(a-2w)+$$

Replace the differences of even order in square brackets by the differences of odd order using the relations,

$$\Delta^2 f(a-w)=\Delta f(a)-\Delta f(a-w)$$
$$\Delta^4 f(a-2w)=\Delta^3 f(a-w)-\Delta^3 f(a-2w)$$

Substituting we obtain Stirling s formula in the form

$$f(a+xw)=f(a)+x\left\{\frac{\Delta f(a)+\Delta f(a-w)}{2}\right\}$$

$$+\frac{x}{2!}\Delta^2 f(a-w)$$

$$+\frac{x(x^2-1^2)}{3!}\left\{\frac{\Delta^3 f(a-w)+\Delta^3 f(a-2w)}{2}\right\}$$

$$+\frac{x^2(x^2-1^2)}{4!}\Delta^4 f(a-2w)+$$

Newton Bessel s formula

Gauss formula can be written as

$$f\ a+xw)=\tfrac{1}{2}f(a)+\tfrac{1}{2}f(a)+x\Delta f(a)$$

$$+\frac{x(x-1)}{2!}\left(\tfrac{1}{2}\Delta^2 f(a-w)+\tfrac{1}{2}\Delta^2 f(a-w)\right.$$

$$+\frac{(x+1)\,x\,(x-1)}{6}\Delta^3 f(a-w)+$$

Substituting the values of $\tfrac{1}{2}f(a)$ $\tfrac{1}{2}\Delta^2 f(a-w)$ $\tfrac{1}{2}$ $\Delta^4 f(a-2w)$ from the relations $\Delta f(a)=f(a+w)-f(a)$

$$\Delta^2 f(a-w)=\Delta^2 f(a)-\Delta^3 f(a-w)$$
$$\Delta^4 f(a-2w)=\Delta^4 f(a-w)-\Delta^3 f(a-2w)\quad\ldots$$

In Gauss formula, written above the result is,

$$f(a+xw)=\tfrac{1}{2}\{f(a+w)-\Delta f(a)\}+\tfrac{1}{2}f(a)$$

$$+x\Delta f(a)+\frac{x(x-1)}{2}\tfrac{1}{2}\{\Delta^2 f(a)-\Delta^3 f(a-w)\}$$

$$+\frac{x(x-1)}{2!}\tfrac{1}{2}\Delta^2 f(a-w)+\quad\ldots$$

Rearranging we obtain Bessel's formula as

$$f(a+xw)=\tfrac{1}{2}\{f(a)+f(a+w)\}$$
$$+(x-\tfrac{1}{2})\Delta f(a)+\frac{x(x-1)}{2!}\tfrac{1}{2}\{\Delta^2 f(a-w)+\Delta^3 f(a)\}$$
$$+\frac{x(x-1)}{3!}(x-\tfrac{1}{2})\Delta^3 f(a-w)+\;..$$

Laplace-Everett's formla.

From Gauss formula

$$f(a+xw)=f(a)+x\Delta f(a)+(x_2)\Delta^2 f(a-w)$$
$$+(x+1_3)\Delta^3 f(a-w)+(x+1_4)\Delta^4 f(a-2w)+\cdots\;..$$

Eliminate the differences of odd order, using

$$\Delta f(a)=f(a+w)-f(a)$$
$$\Delta^3 f(a-w)=\Delta^2 f(a)-\Delta^2 f(a-w),\;\ldots\ldots$$

We obtain

$$f(a+xw)=f(a)+x\{f(a+w)-f(a)\}+(x_2)\Delta^2 f(a-w)$$
$$+(x+1_3)[\Delta^2 f(a)-\Delta^2 f(a-w)]+(x+1_4)\Delta^4 f(a-2w)$$

Applying the general result, $^{n+1}C_r={}^{n}C_{r+1}+{}^{n}C_r$

$$\text{i.e.}\quad \binom{n+1}{r+1}=\binom{n}{r+1}+\binom{n}{r}$$

$$f(a+xw)=f(a)\;[1-x]+xf(a+w)$$
$$+(x+1_3)\Delta^2 f(a)-(x_3)\Delta^2 f(a-w)$$
$$+(x+2_5)\Delta^4 f(a-w)-(x+1_5)\Delta^4 f(a-2w)+$$

Transforming the Coefficients of $f(a)$ by $1-x=\eta$, the result can be written in central difference notation as

$$f(a+xu)=u_x=\left[\eta+\frac{\eta(\eta^2-1)}{3!}\delta^2+\quad\right]u_0$$
$$+\left[x+\frac{x(x^2-1)}{3!}\delta^2+\quad.\quad\right]u_1$$

Which is Everett's formula used for interpolating $f(a)$ and $f(a+w)$.

EXERCISE XIV

1 Find the value of $\Delta^3 u$ and ex ess $f(a+6w)$ in terms of (a) and its differences

Ans $u_{x+3} - (3_1)\, u_{x+2} + 3u_{x+1} - u_x$

2 Prove that the $(n+1)$ th d ffe ence of a polynomial of nth degree vanish Represent the funct on $x^4 - 127x^3 + 42x^3 - 30x + 9$ nto factorial and show that the fourth difference s 24

(*M A Aligarh 1943*)

3 Let a b c and d be successive en ries in a difference able corresponding to equidistant arguments show tha when ourth and higher differences are neg ected the entry co responding to the argument half way between the arguments of and c is $\dfrac{9(b+c-(a+d)}{16}$

4 Given log tan 24° = 9 64858 log tan 24°20 = 9 648696 log tan 2440″ = 9 648923 log tan 24°1′ = 9 64892 log tan 24°1′20″ = 9 64903

Find the value of log tan 24°5 *Ans 9 64861*

5 Given log 6 04 = 78103 log 6 041 = 7811 log 042 = 78118 log 6 043 = 78125 log 6 044 = 78132 deter mine the value of log 5 0104 *Ans 78106*

6 Show that (1) $f(abcd) = \dfrac{f\,abc) - f(bcd)}{a-d}$

(2) The divided differences of order n of x^n and that of polynomial of nth degree are constant

7 Establish Lagrange s formula with the help of lternants

8 Given

x	5	11	27	34	42
$f(x)$	23	899	17315	35606	68510

Express $f(x)$ in terms of the powers of $x-3$

(*M A Aligarh 1943*)

Ans $-13 + 2(x-3) + 6(x-3)^2 + (x-3)^3$

9 Show that (1) Grego y Newton s formula is a special case of Newton s formula for unequal intervals

(2) The differential operator D can b* connected with the difference operator Δ

10 Given $\sqrt{12500} = 111\ 803399$ $\sqrt{12510}$ $= 111\ 84811$ $\sqrt{12520} = 111\ 892806$, $\sqrt{12530}$ $= 111\ 937483$ show that $\sqrt{12516} = 111\ 874929$

(*M A Panjab 1943*)

11 Given Sec $88°4' = 61\ 3911$ S-c $89°5' = 62\ 5072$ Sec` $89°6' = 63\ 6646$ Sec $89°7' = 64\ 8657$ show that Sec $89°5'\,40'' = 63\ 274'$

(*M A 1944*)

12 Show that $f(a\ \ a+w\ \ a-w) = \frac{1}{2w^2}\Delta^2 f(a-w)$

$$f(a\ \ a+u\ \ a-w\ \ a+2w) = \frac{1}{3\,!\,w^3}\Delta^3 f(a-w)$$

$$f(a\ \ a-w\ \ a+w\ \ a-2w) = \frac{1}{3\,!\,w^3}\Delta^3 f(a-2w)$$

13 Deduce Gauss Back ward formula from Newton s formula for unequal intervals *i e*

$$f(a-xw) = f(a) - x\Delta f(a-w) + \frac{x(x-1)}{2\,!}\Delta^2 f(a-w) +$$

14 Show how Ne wton s formula and Stirling s formula can be applied to find th- -a'u s of the diff-ren ial coefficients of a giv en function

15 Express the derivatives of $f(x)$ in terms of the divided differences

16 What is sub abulation ? Deri e the formulae for subtabulation with the help of Gregory Newton formula

17 Given the following values obtain the value of $f(x)$ when $x = 42$

x	30	35	40	45	50	55	60
$f(x)$	771	862	1001	1224	1572	2123	2·83

(M. A. 1945) *Ans* *1081 873.*

18. Find the form of the function given that $f(0)=8$ $f(1)=11$ $f(4)=68$, $f(5)=128$ *(M. A 1945).*

Ans $5x^3-9x^2+16x+32$.

19 Given, Sin 25° 40′ = 43313, Sin 25° 40′ 20′ = 43322, Sin 25° 40′ 40″ = 43336 Sin 25° 41′ = 43339, Sin 25° 41′ 20″ = ·43348, find the value of Sin 25° 40′ 30″ by Stirling and Bessel formulæ *Ans.* *43326.*

20 Given, Logarithms of 310 = 2 4914, 320 = 2 5051, 330 = 2 5185, 340 = 2·5315, 350 = 2 544, 360 = 2·5563, apply auy cenral difference formula to find log 349 and log 3375.

Ans *2 5428, 3·5928*

21 Given the following values for $y=\log e^{x}$ at $x=300$, 301, 302, 303, 304, 305, 306, 307, 5·7037, 5 7071, 5 7104, 5 7137, 5 7170, 5 7203, 5 7235, 5 7268, find the values of $\frac{dy}{dx}$ at $x=300$ and 302 *Ans.* *·0033, 00331.*

Hint —Differentiate Newton's formula

22 Given, Sin 25° = 4226, Sin 25° 1′ = ·4229, Sin 25° 2′ = 4231 Subtabulate for Sin 25° 20″ and 40″

Ans *·42271, ·42281.*

23 A root of $x^3+x=3$ lies between 1·2 and 1 3 Find by inverse interpolation its value upto four places of decimals *Ans.* *1·2134.*

24. Show that if $w_x = u_{\frac{x+0}{t}} + u_{\frac{x+1}{t}} + \quad u_{\frac{x+(t-1)}{t}}$,

then the individual value $u_{\frac{x}{t}}$ may be found from the groups of t individual values w_0, w_1, w_2, and their differences by the formula

$$u_{\frac{x}{t}} = \frac{w_0}{t} + (2x-t+1)\frac{\Delta^2 u_0}{2\,1\,t^2}$$

$$+\{3x^2+3x(1-2t)+(1-3t+2t^2)\}\frac{\Delta^3 w_0}{3\,!\,t^3}\cdot$$

neglecting higher differences. (Forsyth)

Hence or otherwise find the value of the quantity for the middle year of the second quinquennium from 44133, 41921, 39387 *Ans* 8387 (*nearly*)

Sol Put $x=7$ and $t=5$ in the formula

25 The population of a country for four consecutive age groups are given by 10 to 14 years (inclusive) 458572, 15—19, 441424 20—24, 423123, 25—29, 402918 use formula in Q 10 or (otherwise) to find the populations of ages between 17 18 and 22 23 years *Ans* *88294, 84640*

26 Prove Euler-Maclaurin formula and apply it to obtain Stirlings approximation to the factorial Explain Bernoulli's numbers (*M A Punjab 1942 & 1943.*)

Obtain a formula for the sum of nth powers of the first k integers

27 Sum the series

(a) $\frac{1}{(201)^2}+\frac{1}{(203)^2}+\frac{1}{(205)^2}+\cdots\frac{1}{(296)^2}$

(b) $\frac{1}{11^3}+\frac{1}{12^3}+\frac{1}{13^3}+$ *ad in f.*

(*M.A Punjab, 1942*)

Ans (*a*) *000833* (*b*) *00452*

(c) Derive Lubbock and Gregory formulae for summation (*M A. 1944*)

28 Explain the method of least squares and describe one of the fundamental methods of solving Normal equations, showing the Mathematical process clearly

(*M A Aligarh, 1942*)

29 Given $4\ 91x-59y=-339\ 8$, $2\ 72x-2\ 7y=-47\ 5$, $05x+32\ 4y=262\ 5$, $-2\ 91x+27\ 7y=152\ 9$, $-4\ 77x+\ 4y=-27\ 9$. Form normal equations and find x and y

(*M A. Aligarh 1941*)

Ans Normal equations are $62\ 73x-382\ 7y=-2096\ 3$ *and* $-382\ 7x+5307\ 3y=32877\ 7$, $x=7\ 81$ *and* $y=6\ 76$ *(See Chapter VIII, for Normal equations etc)*

30 Apply Doolttle s method to solve the normal equations

$$x+3y-2z+0\,u-2v=5\,,\qquad 3x+4y-5z+u-3v=5\ 4\,,$$
$$-2x-5y+3z-2u+2v=-5\,,\qquad y-2z+5u+3v=7\ 5\,,$$
$$-2x-3y+2z+3u+4v=3\ 3$$

Ans $1\ 5,\ -1,\ 4,\ 2\ 7\ \ -1\ 40035$

31 State briefly the characteristic properties of Lexian and Bernoullian distributions Show that the Lexian variance exceeds the Bernoullian one by an amount which increases with n, the number of trials (*M A 1943*)

32 Prove Euler-Maclaurin formula

$$\frac{1}{w}\int_{a}^{a+rw} f(x)\,dx=\tfrac{1}{2}f(a)+f(a+w)+$$

$$+\tfrac{1}{2}f(a+rw)-\frac{w}{12}[f'(a+rw)-f'(a)]+$$

and Compute $\displaystyle\int_{100}^{105}\frac{dx}{x}$

correctly to seven places of decimals

Ans 0487902 (*M A 1943*)

33 Describe the important properties of normal distribution, and derive the equation of the normal frequency curve in the form

$$y=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{x^2}{2\sigma^2}}$$

State characteristic properties of this curve

(*M A 1943 & 1945*)

34 Write a short note on the periodogram anal, and derive the equation of the periodogram

(M.A. 1943 & 1945.

35 Explain the meaning of trigonometric interpolation Determine the co-efficients in the sum

$$a_0+a_1 \cos x+a_2 \cos 2x+ \cdots +a_5 \cos 5x+a_6 \cos 6x$$

$$+b_1 \sin x+b_2 \sin 2x+\ldots +b_5 \sin 5x$$

which takes the given values $u_0, u_1, \cdots u_{11}$ respectively r x takes the values

$$0, \frac{\pi}{6}, \frac{2\pi}{6}, \frac{3\pi}{6}, \quad \frac{11\pi}{6}$$

respectively. [*Hint.*—This is Fourier analysis, n being 12.

(M.A 1943 & 1945

The co-efficients are given by

$$a_0 = \frac{1}{n} \Sigma_{p=0}^{n-1} u_p \quad a_1 = \frac{2}{n} \Sigma_{p=0}^{n-1} u_p \cos \frac{2p\pi}{n},$$

$$a_r = \frac{2}{n} \Sigma_{0}^{n-1} u_p \cos \frac{2p\pi r}{n}.$$

$$b_r = \frac{2}{n} \Sigma_{p=0}^{n-1} u_p \sin \frac{2pr\pi}{n}$$

when $r=\frac{n}{2}$, use $\frac{1}{n}$, instead of $\frac{2}{n}$ in a_r

Thus for $n=12$, $a_1=\frac{1}{6}\Big\{ u_0+u_1 \quad \frac{\sqrt{3}}{2}+u_2 \quad \frac{1}{2}-u_4$

$-\frac{\sqrt{3}}{2} u_5-u_6-u_7 \quad \frac{\sqrt{3}}{2} -u_8 \quad \frac{1}{2}+\frac{1}{2}u_{10}+ \quad \frac{\sqrt{3}}{2} u_{11}, etc. \Big\}$ [Similarly the co-efficients can be worked out for $n=4$ and $n=6$).

36 Write short notes on —

Bivariate normal frequency surface, math expectation; Tests of significance, multiple correl. Correlation Ratio, Regression, Dispersion.

(M A 1945 & 1943)

CHAPTER XVI

INDIAN STATISTICS

(A) A brief summary of Bowley-Robertson inquiry with regard to

(1) Organisation of Statistics in India

(2) Measurement of National income

(3) Census of Production

(B) List of statistical publications in India will be dealt with in this Chapter

(A) Bowley-Robertson report, entitled 'A scheme for an Economic Census of India' 1934) deals with the fundamental points mentioned above

We shall take up briefly the recommendations of the Bowley Robertson Committee one by one

I —Organisation of Statistics — A permanent economic staff consisting of four members, one being the Director of Statistics should be established directly attached to the Ecoromic Committee of the Governor General's Executive Council, for the organisation of the whole work of economic intelligence

The duties of the Director of Statistics should include (*i*) conduct of the census of population (*ii*) Conduct of the census production, (ii) Co ordination of the central and provincial statistics The census of production should be quinquennial (aftrr 5 years) and while the main census of population continues to be decennial a supplementary census mainly devoted to numbers, age sex, and occupation of people should be taken in the middle of the decennium

There should be in each major province a whole time statistician, as nearly independent of departmental control as administrative requirements permit but making his services avaible to all departments The Director of Statistics should as far as possible, have contact with the Statistics of the provinces to promote uniformity in the provincial statistics andt thus facilitating their assembly into all-India totals

It may be pointed out that following the Coommittee's major departments of the central and provincial Governments

II —Measurement of Notionol Income—The authors remark that the materials for estimating the national income in India are very defective and thus they made several practical proposals for the measurement of the total national income in India

' The national income according to the Committee is the money measure of the aggregate of goods and services accruing to the inhabitants of a country during a year, including net increments to or excluding net decrements from their individual or collective wealth

Two methods of calculation of the national income have been pointed, first is an evaluation of goods and services accruing and second is a summation of individual incomes

The first method is the ' census of production , method and the second is the ' census of income ' method

Both the methods may be employed for the purpose, but special caution in combining the results of the two may be necessary especially in the case of India

The census-of-production method involves,

1 Evaluating the net output of agriculture, mining, industry and other productive enterprises at the point of production Double counting (e g counting both the out put of wheat and the labour of the cattle employed in raising it) should be avoided

All that part of the product of agriculture etc , which is consumed by the producer or bartered locally for the services of workers, should be valued, the price prevailing at the point of production to be counted

2 Adding the value which the transporting and merchanting agencies impart to home produced goods and to imports

3 Adding excise duties on home produced goods and custom duties on imports, in order to secure the aggregate of exchange values to the consumer.

4. Adding the value of imports including gold and silver

5 Deducting the value of exports including gold and silver

6 Adding the value of personal services of all kinds

7 Adding the annual rental value of houses, whether rented or occupied by the owners

8 Adding the increments or deducting the decrements of bank balances and securities abroad whether, by individuals or by governments, similarly, deducting the increments or adding the decrements in the holdings of balances and securities in the country by residents abroad.

9 Deducting the value of goods, whether home produced or imported, which are used for the purpose of keeping intact the fixed Capital, raw meterials or finished commodities

The method of census of production (or products) described above is more fundamental of the two methods of evaluting the national income

Certain precautions would have to be observed so that the result of the census of-income method may tally with those of the first

The second method consists of the summation of individual incomes.

Bowley Robbertson make a distinction of rural income and urban income for India

For *rural income* they recommend an estimate of the quantity and value of all goods and services which arise from the land or rendered in the villages, by the method of intensive surveys in selected villages

For urbon income, they recommend *surveys of the larger town* based on a sample inquiry o fthe personnel and occupations of families, and an estimate of their incomes by personal statements and by investigation of wages and salaries prevailing in the towns. For incomes over at least Rs. 2,000, income tax statistics can be of valuable help They have recommended an intermediate urban population census.

These three inquiries would be supplemented by a census of production applied to factories using power, mines and some other industries

All the investigations should be extended to the Indian States so far as they are willing and able to coeperate For areas not so covered, estimates will be necessary by the use of agricultural statistics

Rural sample Surveys—The statistical method for selecting the villages for intensive survey is that of random sampling whi h may be applied as —Prepare a complete list of all the villages in a province, arrange them in geographical order of districts (or in an order that corresponds to various types of cultivation), decide on the number of villages *to be investigated*, and finally, starting *from* a random number, mark the requirsd number of villàges all nearly equally spaced For example, in a province having 110 000 villages, it is decided to investigate 300, villages

The first mark may be placed on any arbitrary number, say 5th village, on the list, the next mark will be on $\left(5+\frac{110000}{300}\right)$ *i e* on 371 village and so on

Thus every unit in the list shall have a chance of being included for inquiry

When the villages have thus been selected, no other should be substituted

The report gives the following table which indicates the proposed minimum number of villages for investigation in each province

Provirce	Number of villages	Number in Sample
Bengal	86,000	250
Bihar and Orissa	83,000	300
Bombay	21,000	200
Certal Provinces	40,000	200
Madras	51 000	200
Punjab	35,000	200
U. P	106,000	300

To make the total for British India some estimates should be made for Assam, N W F Province, tea plantations of Bengal, areas of Bengal where coal mining is important and the areas effected by earthquake and not fully resettled till the period of inquiry

For conducting the investigations and sample surveys, trained investigators and qualified statisticians should be appointed in order to obtain satisfactory bnd reliable results

The investigator should live in the village for a year or so

The work of the investigators may be suppervised by senior investigators

The entire survey should be under the control of the Director of Statistics at the the centre, through the Provincial statistician

The necessary schedules and Questionnaires should be carefully and briefly drawn considering the local circumstances, and they must have local terms of measures weights etc

Although the main inquiry is to be directed to income, production, consumption and allied topics, the investigator may collect information regarding health, co operation debt etc of the people of the village concerned

Urban Surveys—For urban income, as mentioned before, surveys of the larger towns have been recommended Random sampling is not recomended

As most of the larger towns or cit es have University Centre or Colleges, so surveys of such towns should first be conducted Later on, on similar lines, other towns are to be investigated

For the organisation of *University city surveys,* a central committee should be appointed to draw up an outline schedule of enquiry, to advise and present a report on the whole subject at the end In every University and College there will be the Economics Department to help in the conduct of the economic investigation of the towns The staff and the students, backed by official and monetary help can easily undertake the work of inquiry The post graduate students having knowledge of or qualification *in statistics* would prove of good help

In this way the students will also have practical training The Education Department of the Province will also be of great assistance in this matter

When intensive surveys of University and College towns are completed surveys of other towns are to be undertaken and some of the more efficient investigators in the University city surveys be appointed for carrying on the work

An *occupational census* is almost essentail For this census enquiries should be made about current rates of earnings and wages estimated over the years and allowing for seasonal fluctuations

Ea h industry and occupation of any importance must be included and workers in industries clerks municipal and railway employees tonga drivers and all others working for salaries or wages or making petty profits should be given the r due importance The method of payment may also be recorded

An accurate list of houses or tenements should be prepared Big towns containing say nearly 150 000 houses should be subdivided into five units, each having nearly 30 000 houses About 1 000 houses selected in each unit at random should be visited by the investigator and no house is to be substituted

The visitor should have friendly relations with the residents in order to obtain reliable information about numbers sex age and occupation of the *family groups* Schedules and Questionnaires should be filled in immediately after and not dur ng the visit Repeated visits may be essential to collect correct data

The totals should in case of doubt be given varying within a certain range and not as an exact number

All existing data relating to the subject of the survey should be carefully studied and all persons official and non official interested in or concerned with the collection of data should be cousulted with a view to have a reliable and serviceeable data

III —Census of Praduction —The census of production would be imposed by special Act of the Legislative Assembly at the Centre, making the communications of facts demanded compulsory The census would be conducted by the Director of Statistic with the co-operation of the Departments of Industries and Labour.

Industries employing 20 or more persons and using mechanical power, some sm ll workshops, and also some large non mechanical establishments such as brick-making and carpet making industries should be investigated. Railways and all establishments under the mines Act should also be dealt with.

Since the progress of factory industry is, to a certain extent, at the cost of cottage industry, it will be of great value if the two are brought in statistical relation with each other, and if, some annual data about them could be obtained, it will show their relative increase or decrease

For the purpcse of the census, it is essential that the questionnaires be simple and adapted to Indian environments The essential facts to be elicited are aggregate value of the sales and the aggregate cost of materials for each factory The differance will approximately indicate the national in come accoruing to the factory and when all the factories are taken into account the aggregate differences minus depreciation of plant and change in the value of materials and finished goods will measure the contribution to the national income of the industry

Details can also be obtained of the amounts and values of different commodities produced, and of material purchased and power used

The classification of the produc should be the same as that of exports and imports The employees snould be classed as salaried persons and wage earners, young and adult with an exact statement of the age-division between males and females

To get an average for the year and also as an indication of seasonal variations, it is best to obtain the details of the employees for one week in each month of tl year.

The investigators will face opposition and difficu but with periodic repetition of the census, they wil automatically disappear

B.—List of important statistical publications in India

I Publications of the Department of Commercia Intelligence and Statistics, Government of India.

1. Statistical Abstract for British India (annual)

2 Agricultural Statistics of India —

Vol I—British India (annual)

Vol II—Indian States (annual)

3 Statistical Tables relating to Banks in India (ann al

4 Statistical Tables relating to the Co operativ Movement in India (annual)

5 Large Industrial Establishments in India

6. Review of the Trade of India (annual)

7 Indian Trade Journal (weekly)

8 L ve-Stock Statistics India (quinquennial).

9 Monthly statistics of cotton spinning and weav in Indian Mills

10 Monthly Statistics of the production of cert selected Industries of India

11. Accounts relating to the Sea-borne Trade Navigation of British India (monthly)

12. Accounts relating to the Inland (Rail and Ri borne) Trade of India (monthly)

13 Monthly statement of wholesale prices of ce selected articles at various centres in India.

14. Accbunts relating to the Sea-borne Trade Navigation of British India (annual)

15 Estimates of area and yield of principal crops India (annual).

16 Indian tea coal rubber and coffee Statistics published separately) (annual)

17 Joint Stock Companies in British India and in ome Indian States (annual)

18 Crop forecastes of Rice Wheat Cotton, Linseed, sesamum Groundnut Sugar cane Castorseed (Periodically lso published in the Indian Trade Journal)

19 Quinquennial Report on the average yield per cre of Principal crops in India

20 Crop Atlas of India

II —Other Government (Official) Publications

1 Gazette of India (weekly)

2 Gazettes of Provinces (also of States weekly)

3 Labour Gazette Bombay (monthly)

4 Central and Provincial Government s budgets (annual)

5 Administration Report of Provincial Governments annual)

6 Administration report of Railways in India (annual)

7 Censu Reports (for India Provinces and States) ecennial

8 Report of the Controller of Currency (annual)

9 Monthly survey of business conditions in India

10 Guide to current official statistics

11 Working class Family Budgets

12 India Labour Gazette Monthly

13 Statistics of Factories issued by the Labour Deptt overnment of India

14 Nutrition (Food Department)

15 Publication of Imperial Council of Agricultural esearch

III —Non official publications and Research Journals

1 Sankhya Journal of the Indian Statistical Institute, lcutta

2 Journal of the Indian Mathematical Society, Madras.

3 Proceedings of the All India Science Congress, Statistics Section

4 Monthly survey of economic conditions in the Punjab and other publications of the Board of Economic Inquiry, Punjab, Lahore

5. Capital (Calcutta) Weekly

6 Commerce Bombay Weekly

7 Indian Journal of Economics, Allahabad

8 Indian Year Book

9 Wealth of India, by Wadia and Joshi

10 Wealth and Taxable Capacity by Shah and Khambata.

11. The Indian Finance, Calcutta (Weekly)

12 India s National Income by V K R V Rao

13 Industralisation of the Punjab by Shah

14 Eastern Economist (Weekly) New Delhi

15 Proceedings of the National Academy of Science, Allahabad

16. Journal of the Indian Merchant s Chamber of India.

17 Publications of the Reserve Bank

18 A Plan for Economic Development for India by Sir P Thakar Das and others 1944

IV —Reports of Committees and Commissions

1 Report of the Economic Enquiry Committee (1925)

2 Report of the Royal Commission on Indian Agriculture

3 Report of the Taxation Inquiry Committee

4. Industrial Commission Report

5 Report of the Royal Commission on Indian Labour.

6. Banking Inquiry Committee Reports (Central and Provincial)

7 Reports of the Committees and Commissions on ndian Currency and Exchange

8 Industrial surveys in various districts of U P

9 Labour unemployment and Textile Enquiry Committee Reports (Provincial)

10 Tariff Board Reports.

11 Report of Bowley Robertson Committee

12 Food grains Policy Committee Report (1943)

APPENDIX

I—QUESTIONNAIRE FORM

BOARD OF ECONOMIC INQUIRY PUNJAB
LAHORE

Socio Economic Survey of Greater Lahore

Housing Conditions

1 Ward and Locality

2 Mohalla Road Street

3 Lane (if any .. Width of the lane o street

4 House No 5 Name of owner

6 (*a*) Owner s religion and nationality Hindu Muslim, Sikh Indian Christian Parsee Anglo Indian, European others (specify)

(*b*) Owner s domicile (Province or State)

7 Owner s Occupation

8 Kind of dwelling : Bungalow house hut

9 Nature of dwelling Pacca, Kachha temporary structure

10 Year in which built

11 Number of Storeys (excluding underground ac omodation) one two three four five

12 Total height of building in feet from ground level

13 Number of underground rooms (if any)
How are these rooms used ? -- -

14. Total area of land (in marlas or square feet) - -
Area of the open space or courtyard (if any)
How is the open space or courtyard used?

15 Total cubic space (in cubic feet) of the covered living rooms on
Greund floor
1st floor -- --
2nd floor 3rd floor -
4th floor 5th flo r Grand total

16 Total number of families living
(a) owners (b) tenants Total

17 Total number of occupants (exclude visitors servants living outside the building)
(a) males (b) fomales Total

18 Cubic space per person

19 Is there any Electric power connection ? Ye no

20 Source of water supply Inside the house outside it
If nside, whether Municipal tap, private tubewe hand pump open well

If hand pump or open well, quality of water Swee Saltish

21 Total number of latrines -- -
No of latrines located on ground floor
first floor 2nd floor top floor

Nature of floor of the latrines How many [illegible]
pacca broken

How often cleaned daily ? Once, twice

How many have flush system ?

How many are combined with bath room ?

22. Drainage outside the house pacca, kachba,
Are the drains connected with the main su drains ? Yes, no

Are there any water troughs (*houdi*) ? Yes no
If so whether kachha pacca
Who cleans them ? Sweeper corporation lorry, not cleaned
Specify if any stagnant water collects anywhere

23 Other nuisances Rats bugs bad smell

24 How many times a year is the house whitewashed? .

25 Is any shop attached with the house ? Yes no
If so nature of the business carried on — —
Cubic space occupied by the shop — —
Is the house owner himself the business man ? Yes no
If not rent paid by the shopkeeper .. — —

26 General condition of the house — —
— — — —

For each Family

ouse No . Family No —

1 Family Owner, Tenant

2 (*a*) Religion and nationality Hindu Muslim Sikh, Christian Parsee Indian Christian Anglo Indian, European Others (specify)
(*b*) Domicile (Province or State) ..

3 Since when living there . —

4 Occupatian of Earners and distance of places of their work from the house —
(*a*)
(*b*)
(*c*)

. —

5 Numbers actually living (exclude servants livin outside)—

	Male	*Females*	*Total*
Adults			
Children (5 15 years)			
Babies (below 5 years			
Grand total			

	Males	*Females*	*Total*
6 Literates (Above 5 years)			
7 Married			
Widowed			
Unmarried			

8 No of living rooms

(*a*) how many are completely dark ?

(*b*) how many are well ventilated ?

(¼th of the base area of the room opening into external air)

9 No of separate kitchens Nil one two

Kitchens having chimneys Nil one two

Bath rooms Nil one, two

Godowns Nil one, two

Garage Nil one two

Latrines Nil one, two

10 If no drinking water arrangement inside the house state distance of the source of supply of water in yards

11 Lighting arrangement Electricity kerosene o rapeseed oil

12 Fuel used firewood, charcoal, soft coke, dung saw dust electricity

13 Where does the family sleep in summer ? top fl

inside the room, verandah, outside the house, open space

14 No of separate servants' quarters (if any) ——
Total cubic space (in cubic feet) — —
No of persons living — —

15 No of domestic animals, if any, cows, — —
buffaloes goats sheep — .
horses dogs — poultry — —

16 Is there any separate accommodation for domestic animals ? ..Yes, no.

17 Approximate monthly income of the family , —

18 If tenant monthly rent paid — . —

Filled in by . ...

Date ... .

II —INTERPRETATION OF DATA

Statistical methods are liable to misuse either deliberately or unintentionally.

When the methods are not correctly applied statistics are not to be blamed for their unreliable characters, or for wrong interpretation, *but* the persons who are handling them without having good knowledge of the science of statistics Only qualified persons in statistics should take up the analysis and interpretation of the statistical data Interpretation means drawing inferences from an analytical study of the collected data

In any inductive reasoning statistical methods play a prominent part Before giving any judgment and in drawing conclusions and inferences care must be taken to see that the data are *sufficient, homogeneous and comparable*, and the effects of all the other disturbing factors have been fully taken into account.

Statistical data are often interpreted wrongly due to false generalisation For example, statistics with regard to the increase in quantity and value of imported goods are quoted to justify the conclusion that people are in a prosperous

This conclusion would be valid only when the consumption of indigeneous goods is not decreasing to a greater extent. Increase in the consumption of articles of luxury would show general prosperity only when majority of the people get benefit.

Sometimes *mistakees are made in wrongly interpreting* averages, index numbers, co efficient of correlation and co-efficient of association

In short, statistics should be carefully collected unbiassed errors, statistical methods should be skilfully intelligently applied, by tatisticians, and tested according to various tests of significance, in order to have reliable analys and interpretation of data

III —MATHEMATICAL PROOFS OF THEOREMS ON PROBABILITY AND MOMENTS

In chapter XI the statements of the Addition and plication theorems of Probability were given Here we give simple proofs for them

Let ths main event E, fall in n groups of subsi events of which only one can happen in a single trial but which any cne will bring the event E Let t denote the number of equally likely cases Of the possible cases f be in favour of the event The favourable group of ca may be divided into n subgroups of which f_1 are favourabl for the happening of the subsidiary event E_1, f_2 in favour E_2, f_n in favour of E_n Therefore the probability p the whole event E

$$= \frac{f}{t} = \frac{f_1+f_2+f_3+\cdots f_n}{t} = \frac{f_1}{t} + \frac{f_2}{t} +$$

$$= p_1+p_2+p_3+ \quad +p_n$$

which proves the *Addition Theorem.*

Multiplication Theorem.—Let the number of possib cases, for the whole event E be t, for E_1 be t_1,———for E_n be

Each of the t_1 possible cases corresponding to the E_1 may occur simultaneously with each of the t_2

corresponding to the event E_2 Thus there will be altogether $t_1 \times t_2$ cases falling on the events E_1 and E_2 at the same time

Continuing the reasoning, the total number cf equally likely cases resulting from the simultaneous occurrence of the events E_1, E_2 will be $t_1 \times t_2 \times t_3 \times \ldots \times t_n$

If f denote the favourable cases for the whole event E, $f_1, f_2, \ldots f_n$ the favourable cases for E_1, E_2, then following the above reasoning the probability for the whole

event E is $p = \frac{f}{t} = \frac{f_1 \times f_2 \times f_2 \times \ldots \ldots f_n}{t_1 \times t_2 \times t_3 \times \ldots \quad t_n}$

$$= \frac{f_1}{t_1} \times \frac{f_2}{t_2} \times \frac{f_3}{t_3} \times \quad \ldots \frac{f_n}{t_n}$$

$$= p_1 \times p_2 \times p_3 \quad \ldots \times p_n$$

which proves the theorem

The theorem holds for dependent as well as Independent events.

To determine the mean axd variance for ths binomial $(q+p)^n$

From the expansion of $(q+p)^n$ the frequencies corresponding to the number of successes 0, 1, 2 . . ., n

are the terms, q^n, $nq^{n-1}p$, $\frac{n(n-1)}{2} q^{n-2}p^2 \ldots \quad p^n$.

Taking O as the Provismal mean for the series 0, 1, 2, ..n of successes, the deviations (D) well be D = 0. 1 $\quad n$

and $f = q^n, nq^{n-1}p \ldots \quad \ldots p^n$.

$D = nq^{n-1}p + n(n-1)q^{n-2}p^2 + \quad \ldots np^n$.

$= np[q^{n-1} + (n-1)\, q^{n-2}p + \ldots \ldots p^{n-1}]$

$= np[q+p]^{n-1} = np$ since $q+p = 1$.

$\Sigma f = q^n +$

$= 1$

The Arithmetic mean $= O + \frac{\Sigma f\ D}{\Sigma f}$

$= \frac{np}{1} = np$ since sum of frequencies is $(g+p)^n = 1$.

To find the standard deviation and variance let us find the value of $\frac{\Sigma f D^2}{\Sigma f}$

$$\Sigma f\ D^2 = O + n\ q^{n-1}p + 2n(n-1)\ q^{n-2}p^2 + \quad n^2p^n$$

$$= np\Big[q^{n\ \ 1} + 2(n-1)q^{n-2}p + \frac{3(n-1)(n-2)}{2}q^{n-3}p^2$$

$$+ \quad np^{n-1}\Big]$$

$$np\Big[\Big\{\ q^{n-1} + (n-1\ q^{n-2}p + \frac{(n-1)(n-2)}{2}q^{n-3}p^2 \quad + p^{n-1}\Big\}$$

$$+ \Big\{(n-1)q^{n-2}p + \frac{2(n-1)(n-2)}{2}\quad q^{n-3}p^2 + ..$$

$$+ (n-1)\ \ p^{n-1}\Big\}\Big]$$

$$= np[(q+p)^{n-1} + (n-1)p\{q^{n-2} + (n-2)q^{n-3}p) .. + p^{n-2}\}],$$
$$= np[1 + (n-1)p(q+p)^{n-2}]$$
$$= np[1 + p(n-1)] = np + n^2p^2 - np^2$$

Variance is given by the formula

$$\frac{\Sigma f D^2}{\Sigma f} - \left(\ \frac{\Sigma f D}{\Sigma f}\ \right)^2$$
$$= np + n^2p^2 - np^2 - (np)^2$$
$$= np(1-p) = npq$$

and $\quad \sigma = \sqrt{npq}$

Moments In chapter X, the moments about the mean are given in terms of the moments about any arbitrary origin Here we shall establish these relations

If A denote the provisional mean, the moments about any mean are defined by

$$\underset{r}{V} = \frac{1}{n}\Sigma\,(x-A)^r = \frac{1}{n}\quad \Sigma\, D^r$$

If M denote the arithmetic mean, the moments about the Mean are given by $\nu_r = {}^{1}\ \ \Sigma(x-M)^r = \frac{1}{n}\Sigma d^r$.

$V_1 = \frac{1}{n}\Sigma f D$ where n stands for the sum of the frequencies $= \Sigma f$

It is known that

Arith. Mean $= A + \frac{\Sigma f D}{n}$

$\therefore \quad V_1 = M - A.$

$$\mu_1 = \frac{1}{n}\Sigma f d = \frac{1}{n}\left\{f_1(x_1 - M) + f_2(x_2 - M) \quad (+ f_n(x_n - M)\right\}$$

$= \frac{1}{n}\left\{\Sigma f \ x - nM\right\} = 0$, since

$M = \frac{\Sigma f x}{n}.$

Let us establish in general μ_r's in terms of V_r

$$D = x - A = (x - M) + (M - A)$$
$$= d + V_1.$$

$$\mu_r = \frac{\Sigma f d^r}{n} = \frac{1}{n}\left[\Sigma f (D - V_1)^r \right]$$

$$= \frac{1}{n}\left\{\Sigma f (D^r - rD^{r-1}V_1 + \frac{r(r-1)}{2} D^{r-2}V_1^2 - \frac{r(r-1)(r-2)}{3_1} D^{r-3}V_1^3 + \ .(-1)^r V_1^r\right\}$$

$$= \frac{1}{n}\Sigma f D^r - r.\ V_1 \frac{1}{n} \Sigma f D^{r-1} + \frac{r(r-1)}{2}$$

$$V^2_1 \Sigma f D^{r-f} + \cdots\cdots (-1)^r \ V^r{}_1$$

$$= V_r - r \ V_{r-1} V_1 + \frac{r(r-1)}{2} \ V_{r-2} V_1^2 + (-1)^r V_1^r.$$

Putting $r = 1, 2, 3, 4$ .. we express the moments about the mean in terms of the moments about the Provincial Mean.

IV.—Pnnjab University Question-Papers fcr Sta in 1946 Examination. (*Attached*).

CERTIFICATE IN STATISTICS (C ST.) EXAM 1946 PAPER I AND M A. ECONOMICS—PAPER V (b) OPTION (ii)

STATISTICS

Time allowed Three hours
Maximum Marks . 100

Only five questions are to be attempted atleast two of which must be from each of Sections A and B

All questions carry equal marks

SECTION A

1 Suggest a plan for social economic survey of Lahore. Give details

2. Write an essay on the analysis of time series.

3. You are asked to compile a working class cost of living Index for Lahore Suggest a plan Give details

4 How is a population census organised in India? State the methods of obtaining inter-censal year estimates of the population

5 Write notes on the statistical concept of —

(a) Frequency distribution.

(b) Standard Deviation.

(c) Correlation

SECTION B

6 The following table shows the age distribution of married females according to sample census of 1941 in the Baroda State —

Age	*No of mirried females*
0 and above	3
5 ,, ,,	31
10 , ,,	410
15 , ,	1809
20 ,	2446
25 ,, ,	2223
30 ,, ,,	1723
35 , ,,	1292
40 ,, ,,	963
45 , ,,	762
50 , ,,	531
55 ,, ,,	317
60 , ,,	156
65 , ,,	59
70 ,, ,,	37
All ages	12762

Draw a graph showing tle number of married females younger than any given age Hence or otherwise calculate the median age of married females and also the two quartiles, upper and lower

Ans 28 783 . 21'916 , 38 585

7. Fit a stralght line to the following data showing the yield of wheat in bushels per acre from the same plot during 20 years

1855	1856	1857	1858	1859	1860	1861
29.62	32 38	43 75	37 56	30 00	32 62	33 75
1862	1863	1864	1865	1866	1867	1868
43 44	55 56	51 06	44 06	32 50	29 13	47 81
1869	1870	1871	1872	1873	1874	
39 00	45,50	34 41	40 69	35 81	38 19	

Ans 36 375+ 235x (185 4 origin)

8 The correlation Table given below shows for each of 78 towns (1) measures of the amount of over crowding present in a given year and (2) the infant mortality rate in the same year Calculate the co efficient of correlation between over crowding and infant mortality rate

Infant Iortality Rate	*Percentage of population in private families living more than two persons per room*						*Total*
	1 5—4 5	—7 5	—10 5	—13 5	—16.5	—19 5	
36 —	5						5
46	9	1					10
56	10	4	1			1	16
66	4	7	5	2			18
76	2	5	4	1	1		13
86		2	2	2		1	7
96	--	1	2	2	1	1	7
106—116		1	—	1			2
Total	30	21	14	8	2	3	78

Ans '6 5 (approx).

9 Use the method of interpolation to obtain the value of y for $x = 7$ 5 from the following data

x	4	5	6	7	8	9	10
y	941	948	967	1004	1065	1156	1283

Ans. 1031'125.

M A (MATHEMATICS,—PAPERS IV, V, VI (OPTION F)

STATISTICS

Time allowed Three hours

Maximum Marks 100

N.B —Not more than *nine* questions should be attempted All questions carry equal marks, and *six* carry full marks. Greater credit will be given to complete questions correctly answered than to a proportionate number of fragmentary answers

1 Assuming the Gregory Newton Formula of Interpolation, obtain the expressions for the first two differential co-efficients of the function $f(x)$ for the value 'a' of its argument, in terms of the differences.

Given the following data, compute the first two differential co efficients of the function 'y' corresponding to the argument $x=11$

x	y
2	1,08,243
5	1,21,551
9	1,41,158
13	1,63,047
15	1,74,901

2 Obtain the expression for the Euler Maclaurin Formula for the summation of series

Apply the formula to sum the following series to infinity

$$\frac{1}{101^2}+\frac{1}{103^2}+\frac{1}{105^2}+\frac{1}{107^2}+\dots$$

3 Explain the method of forming the Normal Equations

for a set of variables in which the number of equations given is greater than the number of unknowns

Discuss the method of solving these equations by the method of determinants.

4 Define ‘ probability ’ and explain the terms, ‘ Mutually Exclusive ’, and ‘ Mutually Independent ’ as applied to events.

Given n independent events with respective prpbabilitie of occurrence $p_1, p_2 \quad p_n$, prove that one of the probability of at least one of the events happening is

$$\Sigma p_1 - \Sigma p_1 p_2 + \Sigma p_1 p_2 p_3 - \quad \dots$$

This sides of a rectangle are chosen at random, each being less than a given length ‘ a , all such lengths being equally likely Find the chance that the diagonal is less than ‘ a ’.

5 The following table gives the monthly average production of boots and shoes in U S A Fit a curve of the the form $a+bx+cx^2$ to this data

Years	1923	1924	1925	1926	1927	1928	1929	1930	1931	1932	193
Average production (in millions) ..	29 3	26 1	27 0	27 5	28 6	29 2	30 1	25 4	26'4	26 1	29

6 Assuming the conditions of simple Sampling, how do you test the significance of the difference between the ‘ values of Arithmetic Mean ’ and ‘ Standard Deviation ’ obtained from a sample with those of the total population.

In studying the problem of density of population per house, from a population of 1,00,000 houses, a random sample of 1,000 was selected and the following results were obtained

NUMBER OE PERSONS PER HOUSE

	1	2	3	4	5	6	7	8	9	10
Whole population/100	56	216	243	199	124	75	44	22	11	6
Sample	54	225	237	193	121	79	41	27	10	8

Compute the values of Arithmetic Mean, and Standard Deviation of the number of persons per house, both for the whole population as well as for the sample Are the values of Estimates of these two from the sample, significantly different from those of the population ?

7 The number of males in each of 106 eight pig litters was found and they are given by the following frequency distribution —

Number of males per litter →	0	1	2	3	4	5	6	7	8	Total
Frequency	0	5	9	22	25	26	14	4	1	106

Assuming that the probability of an animal being male or female is even (*i e* $p=q=\frac{1}{2}$), *and the frequency* distribution follows the Binomial law, calculate the expected frequencies of the nine classes Find also the values of ψ^2 to test the goodness of fit

8. If x_1 is the dependent variable, and x_2 and x_3 the two independent variables, obtain the regression eq of x_1 in terms of x_2 and x_3.

Give the following values of Arithmetic Mean, Deviation and Co-efficient of Correlation of 740 sets of values find the regression equation of x_1 in terms of x_2 and x_3

$\bar{x}_1 = 28\ 02$	$\sigma_1 = 4{\cdot}42$	$r_{12} = 0{\cdot}80$
$\bar{x}_2 = 4{\cdot}91$	$\sigma_2 = 1\ 10$	$r_{13} = -0{\cdot}40$
$x_3 = 594$	$\sigma_3 = \ 85$	$r_{23} = -0\ 56$

9. x and y are two correlated variables, measured fr their respective arithmetic means. If the standard devi of each is unity and the co efficient of correlation b the two is r, for what values of θ are the two vari $X = x \cos\theta + y \sin\theta$, $Y = x \sin\theta + y \cos\theta$, uncorrelated What are the values of the standard deviations of the variables X and Y ?

10. The following table gives the results of exper on four varieties of a crop in 5 blocks of plots :—

		Block				
		1	2	3	4	5
	A	32	34	33	35	37
Variety	B	34	33	36	37	35
	C	31	34	35	32	36
	D	29	26	30	28	29

Prepare the table of analysis of variance to test the ignificarce of difference between the yields of the four rieties

11 Write short notes on *any four* of the following —

(*a*) Poisson's Distribution. (*b*) Lexian Ratio (*c*) o-efficient of contingency. (*d*) Sheppard's Corrections (*e*) lethod of Moments (*f*) Fourier's analysis and its applica-on to time series

B COM 1946

STATISTICS

Time allowed three hours.

Maximum Marks 100-

Answer five questions only of which at least two must e from Group A and two from Group B All questions rry equal marks

GROUP A

1 Define Statistics and point out the main difficulties at a statistician has to face as compared with a physicist r chemist

How will you classify a given commercial data ?

2. It is required to estimate the total consumption of od grains in the Punjab for enforcing a scheme of food ationing. What statistical data should be collected for this rpose and how ?

3 How would you use the method of random sampling n making an economic survey of villages in the Punjab ?

4 What is the importance of the census of and that of production ?

How will you organise those censuses in your ᴗ Province ?

5 Write notes on three of the following giving example

Probability Mode Tabulation Moving A Index numbers Regression

GROUP B

6 The following table gives five yearly percentage in Bombay Presidency under cotton and under food Calculate the co efficient of correlation between the a under cotton and the area under food crops —

Year	*Percentage area under cotton*	*Percentage under food*
1908	38 5	52 7
1909	38 5	52·3
1910	38 8	53 0
1911	37 8	53 5
1912	39 1	52 5
1913	39 5	52·3
1914	38 0	54 9
1915	38 4	54 3
1916	38 8	53 2
1917	39 2	52 6

Ans —

7 The following table gives the population of Lucknow the time of the previous censuses —

1891	2,64,953
1901	2,56 239
1911	2,32,332
1921	2,17,167
1931	2,51,097

Estimates the population of Lucknow for 1916

Ans 221520

8. The following table gives the d tail of monthly expenure of three families —

'ms of Ex	*Family A*		*Family B*		*Family C*	
penditure	Rs	a	Rs	a	Rs	a
ıod	12	0	30	0	90	0
othing	2	0	7	0	35	0
oure-rent	2	0	8	0	40	0
lucation	1	8	3	0	12	0
tigation	1	0	5	0	40	0
ınventional necessity	0	8	3	0	60	0
ıscellaneous	1	0	4	0	23	0

Represent the above figures by a suitable diagram hich family is spending the money most wisely ?

9 (*a*) Following are the group index numbers, and e group weights of an average working class family's ıdget Construct the cost of living index number by signing the given weights

Group	*Index number for January 1943*	*Weigh*
Food	152	48
Fuel and lighting	110	6
Clothing	130	8
House-rent	100	12
Miscellaneous	90	15

(*b*) Calculate the variance for the given index nu ' in (*a*). *Ans 129 73 , 49*

B A HONS.

ECONOMICS PAPER III OPTION (III) STATISTI

Time allowed : three hours

Maximum Marks 60.

Attempt five questions, atleast two being from C and two from Group B. All questions carry equal marks.

GROUP A.

1. ' The application of statistical methods is extensi But their application in economic and social life man to day most intimately ' (Dr. Sir Manabar Lal).

Elucidate with illustrations the above statement on utility of statistical methods in the present day and social conditions

2. What is the importance of graphic charts business statistics ? What are the various types of diag charts and graphs commonly used ? What precau should be taken in using pictorial or popular presentations

3 What do you understand by skewness ? W the various methods of its measurement ? Illustrate answer by suitable example.

4 What is the use of a cost of living index number? How is it constructed? What are its drawbacks?

5 Write explanatory notes on any three of the following —

(i) Sampling

(ii) Statistical errors

(iii) Lorenz curve

(iv) Seasonal fluctuatione

(v) Chain base index numbers

Group B

6. Present the data given in the following paragraph in the form of a table, so as to bring out clearly all the facts, indicating the source and bearing a suitable title

According to tthe Census of Manufacturers Report' 1945 the John Smith Manufacturing Company employed 400 Non-union and 1250 union employees in 1941. Of these 220 were females of which 140 were non-union. In 1942, the number of union employees increased to 1475 of which 1300 were males Of the 250 non-union employees 200 were males In 1943, 1700 employees were union numbers and 440 were non union Of all the employees in 1943, 250 were females of which 240 were union members. In 1944, the total number of employees was 2000 of which one per cent. were non union. Of all the employees in 1944, 300 were females of which only 5 were non union

7. "Capital" Index of Indian cotton consumption f, January 1944 to May 1945 is given below :—

1944	*Index*	*1944*	*Index*
January	157·5	October	154·2
February	156·1	November	165·9
March	158·9	December	162·6
April	148·1	*1945*	
May	153·3	January	163·1
June	161·7	February	148·1
July	157·5	March	174·3
August	160·3	April	158·9
September	161·2	May	165·9

Represent the above data in the form of a ' _ and indicate the trend based on three-monthly mov' average

8 Compute the Standard Deviation and the co effic of variation from the following data of monthly wages p in a cotton factory —

Wage grades Rs	Number of employees
15—25	7
25—35	102
35—45	111
45—55	360
55—65	159
65—75	33
75—85	13
85—95	11
95—105	0
105—115	4
Total	600

9 From the following record of marks obtained in Economics by a batch of 55 students, indicate the value of the median and the modal marks

12, 17, 18, 20, 20, 24, 25, 28, 30, 30, 33, 33,
33, 33, 33 33, 34, 34 35, 35, 36, 37, 38, 40,
40, 40, 42, 44, 45, 45, 48, 48, 48, 48, 48 48,
49 50, 50, 50 51, 52, 53 54, 55, 56, 58, 58,
59, 59 61, 62, 64, 65, 68

CERTIFICATE IN STATISTICS 1946

PAPER II.

Time allowed three hours
Maximum Marks 100

Atrempt five questions only at least two from each Group. All questions are of equal value

GROUP A

1 Write a note explaining the various uses of Fisher's Z statistics

Or

'*t*' tests

2 Explain the importance of 'replication', 'randomisation' and 'local control' in agricultural field experiments, and mention some of the devices by which local control is achieved

3 Write a short essay on the use of 'Control charts' or on Official statistics in India

4. It is required to determine the percentage of literates in your district Give any sample survey scheme to obtain he desired information

5 Define ‘ multiple ’and partial correlation, and explain with illustrations, the use of these statistical concepts.

GROUP B

6 Find by interpolation the missing value in the following table —

Degrees of freemod	*One per cent value of E*
3	5 841
4	4 604
5	4·032
6	.
7	3 499
8	3 355
9	3 250

7 The following table gives the frequency distrit of expenditure on food per family per month among working class families in two localities Find the mean and standard deviation at both places, and test whether there is any real difference in the expenditure on food at these two places.

Expenditure in Rs per month	*Number of Families*	
	Place A	*Place B*
3—6	28	39
6—9	292	284
9—12	389	401
12—15	212	202
15—18	59	48
18—21	18	21
21—24	2	5

8. Calculate the first four moments for the frequency distribution —

x	89	86	74	65	64	63	66	67	72	79
f	92	91	84	75	73	72	71	75	78	84

9 From the following table showing the number of plants having certain characters, test the hypothesis that the flower colour is independent of flatness of leaf

	Flat Leaves	*Curled Service*	*Total*
White flowers	99	36	135
Red flowers	20	5	25
Total	119	41	160

You may use the following table giving the value of ψ^2 (chi-square) for one degree of freedom, for different values of P.

P	99	·95	·90	·50	·10	·05
ψ^2	000157	·00393	0158	·455	2 706	3 841
P	01					
ψ^2	6 635					

10. Set up a table of analysis of variance for :—

Plots	*Varieties*			
-	*a*	*b*	*c*	*d*
1	200	230	250	300
2	190	270	309	270
3	240	150	145	180

TABLES OF LOGARITHMS, ANTI-LOGARITHMS, SQUARES, SQUARE ROOTS AND RECIPROCALS

The logarithm of a number consists of (1) integral part known as characteristic (2) Decimal part, known as Mantissa

The table of Logarithms gives the Mantisas upto four places of decimals, for numbers of three digits, for ready reference of the students To find the Mantissa of any given number, take the number approximately to three digits and the table will give the approximate value. Mantissa of a number will be the same irrespective of the position of the decimal point in it

If more accuracy is required in the results, then five figure tables or seven figure tables should be used

The characteristics are to be found as

(1) When the given number is greater than 1, the characteristic will be positive and equal to $n-1$, where n is the number of significant digits before the decimal point The characteristic in 514 98 is 2 and log 514'98 =2+'7118=2 7118 nearly

(2) When the given number is less than 1, the characteristic is negative and is greater by one than the number of zeros which follow the decimal point The characteristic of 0034 is 3 (negative) and is written as $\bar{3}$.

For Anti-logarithms the reverse of (1) and (2) are to be utilised.

The number, from the Antilog tables, whose log is 1'6928 is 49'32.

Tables of Squares, etc., are for numbers up to 100. For higher calculations tables such as Barlow's Tables for squares, etc which gives for integers up to 12500 may be consulted. Calculating Machines like Facit, Brunsviga can also be used for rapid and heavy caloplations.

LOGARITHMS

Log 1 = 0 , log 2 = 301, log 3 = 4771 log 4 = 6021 log 5 = 699
log 6 = 7782 = log 7 = 8451 log 8 = 9031 log 9 = 9542

	0	1	2	3	4	5	6	7	8	9
10	0000	0043	0086	0128	0170	0212	0253	0294	0334	0374
11	0414	0453	0492	0531	0569	0607	0645	0682	0719	0755
12	0792	0828	0864	0899	0934	0969	1004	1038	1072	1106
13	1139	1173	1206	1239	1271	1303	1335	1367	1399	1430
14	1461	1492	1523	1553	1584	1614	1644	1673	1703	1732
15	1761	1790	1818	1847	1875	1903	1931	1959	1987	2014
16	2041	2068	2095	2122	2148	2175	2201	2227	2253	2279
17	2304	2330	2355	2380	2405	2430	2455	2480	2504	2529
18	2553	2577	2601	2625	2648	2672	2695	2718	2742	2765
19	2788	2810	2833	2856	2878	2900	2923	2945	2967	2989
20	3010	3032	3054	3075	3096	3118	3139	3160	3181	3201
21	3222	3243	3263	3284	3304	3224	3345	3365	3385	3404
22	3424	3444	3464	3483	3502	3522	3541	3560	3579	3598
23	3617	3636	3655	3674	3692	3711	3729	3747	3766	3784

24	3802	3820	3838	3855	3874	3892	3909	3927	3945	3962
25 →	3979	3997	4014	4031	4048	4065	4082	4099	4116	4133
26	4150	4166	4183	4200	4216	4232	4249	4265	4281	4298
27	4314	4330	4346	4362	4378	4393	4409	4425	4440	4456
28	4472	4487	4502	4518	4533	4548	4564	4579	4594	4609
29	4624	4639	5654	4669	4683	4698	4713	4728	4742	4757
30	4771	1786	4800	4814	4829	4843	4857	4871	4886	4900
31	4914	4928	4942	4955	4969	4983	4997	5011	5024	5038
32	5051	5065	5079	5092	5105	5119	5132	5145	5159	5172
33	5185	5198	5211	5224	5237	5250	5263	5276	5289	5302
34	5315	5328	5310	5353	5366	5378	5391	5403	5416	5428
35	5441	5453	5465	5478	5490	5502	5514	5527	5539	5551
36	5563	5575	5587	5599	5611	5623	5635	5647	5658	5670
37	5682	5694	5705	5717	5729	5740	5752	5763	5775	5786
38	5798	5809	5821	5832	5843	5855	5366	5877	5888	5899
39	5911	5922	5933	5944	5955	5966	5977	5988	5999	6010
40	6021	6031	6012	6053	6064	6075	6085	6096	6107	6117
41	6128	6138	6149	6160	6170	6180	6191	6201	6212	6222

LOGARITHMS—(*contd.*)

	0	1	2	3	4	5	6	7	8	9
42	6232	6243	6263	6263	6274	6284	6294	6304	6314	6325
43	6335	6345	6365	6365	6375	6385	6395	6405	6415	6425
44	6435	6444	6464	6464	6474	6484	6493	6503	6513	6522
45	6532	6542	6551	6561	6571	6580	6590	6599	6609	6618
46	6628	6637	6646	6656	6665	6675	6684	6693	6702	6712
47	6721	730	6739	6749	6758	6767	6776	6785	6794	6803
48	6812	6821	6830	6839	6848	6857	6866	6875	6884	6893
49	6902	6911	6920	6928	6937	6946	6955	6964	6972	6981
50	6990	6998	7007	7016	7024	7033	7042	7050	7059	7067
51	7076	7084	7093	7101	7110	7118	7126	7135	7143	7152
52	7160	7168	7177	7185	7193	7002	7210	7218	7226	7235
53	7243	7251	7259	7267	7275	7284	7292	7300	7308	7316
54	7324	7332	7340	7348	7356	7364	7372	7380	7388	7396
55	7404	7412	7419	7427	7435	7443	7451	7459	7466	7474
56	7482	7490	7497	7505	7513	7520	7528	7536	7543	7551

57	7559	7566	7574	7582	7589	7597	7604	7612	7619	7627
58	7634	7642	7649	7657	7664	7672	7679	7686	7694	7701
59	7709	7716	7723	7731	7738	7745	7752	7760	7767	7774
60	7782	7789	7796	7803	7810	7818	7825	7832	7839	7846
61	7853	7860	7868	7875	7882	7889	7896	7903	7910	7917
62	7924	7931	7938	7945	7952	7959	7966	7973	7980	7987
63	7993	8000	8007	8014	8021	8028	8035	8041	8048	8055
64	8062	8069	8075	8082	8089	8096	8102	8109	8116	8122
65	8129	8136	8142	8149	8156	8162	8169	8176	8132	8189
66	8195	8202	8209	8215	8222	8228	8235	8241	8248	8254
67	8261	8267	8274	8280	8287	8293	8299	8306	8312	8319
68	8325	8331	8338	8344	8351	8357	8363	8370	8376	8382
69	8388	8395	8401	8407	8414	8420	8426	8432	8439	8445
70	8451	8457	8463	8470	8476	8482	8488	8494	8500	8506
71	8513	8519	8525	8531	8537	8543	8549	8555	8561	8567
72	8573	8579	8585	8591	8597	8603	8609	8615	8621	8627
73	8633	8639	8645	8651	8657	8663	8669	8675	8681	8686
74	8692	8698	8704	8710	8716	8722	8727	8733	8739	8745

LOGARITHMS—(*concld*)

	0	1	2	3	4	5	6	7	8	9
75	8751	8756	8762	8768	8774	8779	8785	8791	8797	8802
76	8808	8814	8820	8825	8831	8837	8842	8848	8854	8859
77	8865	8871	8876	8882	8887	8893	8899	8904	8910	8915
78	8921	8927	8932	8938	8943	8949	8954	8960	8965	8971
79	8976	8982	8987	8993	8998	9004	9009	9015	9020	9025
80	9031	9036	9042	9047	9053	9058	9063	9069	9074	9079
81	9085	9090	9096	9101	9106	9112	9117	9122	9128	9133
82	9138	9143	9149	9154	9159	9165	9170	9175	9180	9186
83	9191	9196	9201	9206	9212	9217	9222	9227	9232	9238
84	9243	9248	9253	9258	9263	9269	9274	9279	9284	9289
85	9291	9299	9304	9309	9315	9320	9325	9330	9335	9340
86	9345	9350	9355	9360	9365	9370	9375	9380	9385	9390
87	9395	9400	9405	9410	9415	9420	9425	9430	9435	9440
88	9345	9450	9455	9460	9465	9469	9474	9479	9484	9489
89	9494	9499	9504	9509	9513	9518	9523	9528	9533	9538

90	9542	9547	9552	9557	9562	9566	9571	9576	9581	9586
91	9590	9595	9600	9605	9609	9614	9619	9624	9628	9633
92	9638	9643	9647	9652	9657	9661	9666	9671	9675	9680
93	9685	9689	9694	9699	9703	9708	9713	9717	9722	9727
91	9731	9636	9741	9745	9750	9754	9759	9763	9768	9773
95	9777	9782	9786	9791	9795	9800	9805	9809	9814	9818
96	9823	9827	9832	9836	9841	9845	9850	9854	9859	9863
97	9868	9872	9877	9881	9886	9890	9894	9899	9903	9908
98	9912	9917	9921	9926	9930	9934	9839	9943	9948	9952
99	9956	9961	9965	9969	9974	9978	9983	9987	9991	9996

ANTI LOGARITHMS

	0	1	2	3	4	5	6	7	8	9
00	1000	1002	1005	1007	1009	1012	1014	1016	1019	1021
01	1023	1026	1028	1030	1033	1035	1038	1040	1042	1045
02	1047	1050	1052	1054	1057	1059	1062	1064	1067	1069
03	1072	1074	1076	1079	1081	1084	1086	1089	1091	1094
04	1096	1099	1102	1104	1107	1109	1112	1114	1117	1119
05	1122	1125	1127	1130	1132	1135	1138	1140	1143	1146
06	1148	1151	1153	1156	1159	1161	1164	1167	1169	1172
07	1175	1178	1180	1183	1186	1189	1191	1194	1197	1199
08	1202	1205	1208	1211	1213	1216	1219	1222	1225	1227
09	1230	1233	1236	1239	1242	1245	1247	1250	1253	1256
10	1259	1262	1265	1268	1271	1274	1276	1279	1282	1285
11	1288	1291	1294	1297	1300	1303	1306	1309	1312	1315
12	1318	1321	1324	1327	1330	1334	1337	1340	1343	1346
13	1349	1352	1355	1358	1361	1365	1368	1371	1374	1377

14	1380	1384	1387	1390	1393	1396	1400	1403	1406	1409
15	1413	1416	1419	1422	1426	1429	1432	1435	1439	1442
16	445	1449	1452	1455	1459	1462	1466	1469	1472	1476
17	1479	1483	1486	1489	1493	1496	1500	1503	1507	1510
18	1514	1517	1521	1524	1528	1531	1535	1538	1542	1545
19	1549	1552	1556	1560	1563	1567	1570	1574	1578	1581
20	1585	1589	1592	1596	1600	1603	1607	1611	1614	1618
21	1622	1626	1629	1633	1637	1641	1644	1648	1652	1656
22	1660	1663	1667	1671	1675	1679	1683	1687	1690	1694
23	1698	1702	1706	1710	1714	1718	1722	1726	1730	1734
24	1738	1742	1746	1750	1754	1758	1762	1766	1770	1774
25	1778	1782	1786	1791	1795	1759	1803	1807	1811	1816
26	1820	1824	1828	1832	1837	1841	1845	1849	1854	1858
27	1862	1866	1871	1875	1879	1884	1888	1892	1897	1901
28	1905	1910	1914	1919	1923	1928	1932	1936	1941	1945
29	1950	1954	1959	1963	1968	1972	1977	1982	1986	1991
30	1995	2000	2004	2009	2014	2018	2023	2028	2032	2037
31	2042	2046	2051	2056	2061	2065	2070	2075	2080	2084
32	2089	2094	2099	2104	2109	2113	2118	2123	2128	2133
	21	3	2 4		2 5	2				2

	0	1	2	3	4	5	6	7	8	9
·34	2188	2193	2198	2203	2208	2213	2218	2223	1228	2234
·35	2239	2244	2249	2254	2259	2265	2270	2275	2280	2286
36	2291	2296	2301	2307	2312	2317	2323	7328	2333	2339
37	2344	2350	2355	2360	2366	2371	2377	2382	2388	2393
·38	2399	2404	2410	2415	2421	2427	2432	2438	2443	2449
39	2455	2460	2466	2472	2477	2483	2489	2495	2500	2506
40	2512	2518	2523	2529	2535	2541	2547	2553	2559	2564
41	2570	2576	2582	2588	2494	2600	2606	2612	2618	2624
42	2630	2636	2642	2649	2655	2661	2667	2573	2679	2685
43	2692	2698	2704	2710	2716	2723	2729	2735	2742	2748
44	2754	2761	2767	2773	2780	2786	2793	2799	2805	2812
45	2818	2825	2831	2838	2844	2851	2858	2864	2871	2877
16	2884	2891	2897	2904	2911	2917	2924	2931	2938	2944
47	2951	2958	2965	2972	2979	2965	2992	2999	3006	3013
48	3020	3027	3034	3041	3048	3055	3062	3069	3076	3083
49	3090	3097	3105	3112	3119	3126	3133	3141	3148	3155

.50	3162	3170	3177	3184	3192	3199	3206	3214	3221	3228
·51	3236	3243	3251	3258	3266	3273	3281	3289	3296	3304
·52	3311	3319	3327	3334	3342	3350	3357	3365	3373	3381
·53	3388	3396	3104	3412	3420	3428	3436	3443	3451	3459
·54	3467	3475	3483	3491	3499	3508	3516	3524	3532	3540
·55	3548	3556	3565	3573	3581	3589	3597	3606	3614	3622
·56	3631	3639	3648	3656	3664	3673	3681	3690	3698	3707
·57	3715	3724	3733	3741	3750	3758	3767	3776	3784	3793
·58	3802	3811	3819	3828	3837	3846	3855	3864	3873	3882
59	3890	3899	3908	3917	3926	3936	3945	3954	3963	3972
·60	3981	3990	3999	4009	4018	4027	4036	4046	4055	4064
61	4074	4083	4093	4102	4111	4121	4130	4140	4150	4159
·62	4169	4178	4188	4198	4207	4217	4227	4236	4246	4256
·63	4266	4276	4285	4295	4305	4315	4325	4335	4345	4355
·64	4365	4375	4385	4395	4406	4416	4426	4436	4446	4457
·65	4467	4477	4487	4498	4508	4519	4529	4539	4550	4560
66	4571	4581	4592	4603	4613	4624	4634	4645	4656	4667
67	4677	4688	4699	4710	4721	4732	4742	4753	4764	4775
·68	4786	4797	4808	4819	4831	4842	4853	4864	4875	4887
·69	4898	4909	92	4 3	4	[illegible]	9	497	8	[illegible]

ANTI LOGARITHMS—*(concld.)*

	0	1	2	3	4	5	6	7	8	9
·70	5012	5023	5035	5047	5058	5070	5082	5093	5105	5117
71	5129	5140	5152	5164	5176	5188	5200	5212	5224	5236
·72	5248	5260	5272	5284	5297	5309	5321	5333	5346	5358
·73	5370	5383	5395	5408	5420	5433	5445	5448	5470	5483
·74	5495	5508	5521	5534	5546	5559	5572	5585	5598	5610
·75	5623	5636	5649	5662	5675	5689	5702	5715	5728	5741
·76	5754	5768	5781	5794	5808	5821	5834	5848	5861	5875
77	5888	5902	5916	5929	5943	5957	5970	5984	5998	6012
·78	6026	6039	6053	6067	6081	6095	6109	6124	6138	6152
79	6166	6180	6194	6209	6223	6237	6252	6266	6281	6295
80	6310	6324	6339	6353	6368	6383	6397	6412	6427	6442
·81	6457	6471	6486	6501	6516	6531	6546	6561	6577	6592
·82	6607	6622	6637	6653	6668	6683	6699	6714	6730	6745
·83	6761	6776	6792	6808	6823	6839	6855	6871	6887	6902

.84	6918	6934	6950	6966	6982	6998	7015	7031	7047	7053
85	7079	7096	7112	7129	7145	7161	7178	7194	7211	7228
86	7244	7261	7278	7295	7311	7328	7345	7362	7379	7396
87	7413	7430	7447	7464	7482	7499	7516	7534	7551	7568
88	7586	7603	7621	7638	7656	7674	7691	7709	7727	7745
89	7762	7780	7798	7816	7834	7852	7870	7889	7907	7925
90	7943	7962	7980	7998	8017	8035	8054	8072	8091	8110
91	8128	8147	8166	8185	8204	8222	8241	8260	8279	8299
92	8318	8337	8356	8375	8395	8414	8433	8453	8472	8492
93	8511	8531	8551	8570	8590	8610	8630	8650	8670	8690
94	8710	8730	8750	8770	8790	8810	8831	8851	8871	8892
95	8913	8933	8954	8974	8995	9016	9036	9057	9078	9099
96	9120	9141	9162	9183	9204	9226	9 47	9268	9290	9311
97	9333	9354	9376	9397	9419	9441	9462	9484	9506	9528
98	9550	9572	9594	9616	9638	9661	9683	9705	9727	9750
99	9772	9795	9817	9840	9853	9886	9908	9931	9954	9977

SQUARES SQUARE ROOTS AND RECIPROCALS

No n	Square n^2	Square root $\sqrt{n}$	Reci procal $1/n$	No n	Square n^2	Square root $\sqrt{n}$	Reci procal $1/n$
			o				o
1	1	1 000		26	6 76	5 099	0384
2	4	1 414	5000	27	7 29	5 196	0370
3	9	1 732	3333	28	7 84	5 291	0357
4	16	2 000	2500	29	8 41	5 385	0344
5	25	2 236	20 0	30	9 00	5 477	0333
6	36	2 449	1666	31	9 61	5 567	0322
7	49	2 646	1428	32	10 24	5 656	0312
8	64	2 828	1250	33	10 89	5 744	0303
9	81	3 000	1111	34	11 56	5 830	0294
10	1 00	3 162	1000	35	12 25	5 916	0285
11	1 21	3 316	0909	36	12 96	6 000	0277
12	1 44	3 464	0833	37	13 69	6 082	0270
13	1 69	3 605	0769	38	14 44	6 164	0263
14	1 96	3 741	0714	39	15 21	6 244	0256
15	2 25	3 872	0666	40	16 00	6 324	0250
16	2 56	4 000	0625	41	16 81	6 403	0243
17	2 89	4 123	0588	42	17 64	6 480	0238
18	3 24	4 242	0555	43	18 49	6 557	0232
19	3 61	4 358	0526	44	19 36	6 633	0227
20	4 00	4 472	0500	45	20 25	6 708	0222
21	4 41	4 582	0476	46	21 16	6 782	0217
22	4 84	4 690	0454	47	22 09	6 855	0212
23	5 29	4 795	0434	48	23 04	6 928	0208
24	5 76	4 898	0416	49	24 01	7 000	0204
25	6 25	5 000	0400	50	25 00	7 071	0200

No n	Square n^2	Square root $\sqrt{n}$	Reciprocal $1/n$	No n	Square n^2	Square root $\sqrt{n}$	Reciprocal $1/n$
			o o				o o
51	26 01	7 141	1960	76	57 76	8'717	1315
52	27 04	7 211	1923	77	59 29	8 774	1298
53	28 09	7 280	1886	78	60 84	8'831	1282
54	29 16	7 348	1851	79	62 41	8'888	1265
55	30 25	7'416	1818	80	64 00	8'944	1250
56	31 36	7'483	1785	81	65 61	9 000	1234
57	32 49	7 549	1754	82	67 24	9'055	1219
58	33 64	7 615	1724	83	68 89	9'110	1204
59	34 81	7 681	1694	84	70 56	9'165	1190
60	36 00	7 745	1666	85	72 25	9'219	1176
61	37 21	7 810	1639	86	73 96	9'273	1162
62	38 44	7 874	1612	87	75 69	9 327	1149
63	39 69	7 937	1587	88	77 44	9'380	1135
64	40 96	8 000	1562	89	79 21	9'433	1123
65	42 25	8 062	1538	90	81 00	9'486	1111
66	43 56	8 124	1515	91	82 81	9'539	1098
67	44 89	8 185	1492	92	84 64	9'591	1086
68	46 24	8 246	1470	93	86 49	9 643	1075
69	47 61	8'306	1449	94	88 36	9 695	1063
70	49 00	8'366	1428	95	90 25	9 746	1052
71	50 41	8 426	1408	96	92 16	9'797	1041
72	51 84	8 485	1388	97	94 09	9'848	1030
73	53 29	8 544	1369	98	96 04	9 899	1020
74	54 76	8 602	1351	99	98 01	9 949	1010
75	56 25	8 660	1333	100	10000	10'00	1000

INDEX

BIBLIOGRAPHY.

The author expresses his indebtedness to the ollowing references —

1. Aitken, Statistical Mathematics

2 Annals of Mathametical Statistics, U S A

3 Barlow, Tables of Squares, Cubes, etc

4 Bowley, Elements of Statistics

5 Bowley—Robertson report

6 Brunt, The Combination of Observations

7 Brij Narain, Frequecy Curves, 1944

8 Boddington, Statistics and their application to Commerce

9 Croxton and Cowdon, Applied General Statistics

10 Chambers, Statistical Calculations

11 Colton and Arkin, An outline of Statistical Methohs

12 Census Report 1941

13 Connor, Statistics in Theory and Practice, 1938

14 Dawson, Shepherd, An introduction to the Compution of Statistics

15 Davies and Crowder, Methods of Statistical Analysis

16 Davies and Yarder, Business Statistics

17 Day, Statistical Analysis

18 Dubey and Aggrawal, Elementary Statistics

19 Elderton, Frequency Curves and Correlation

20 Fisher and Yates, Statistical tables for biological, agricultural and medical research

21 Fisher, R A, Statistical Methods for [illegible] workers

22 Fisher A Mathematical Theory of Probability.

23 Forsyth-Math Statistics

24 Ghosh and chowdry Statistics

25 Fry Probability and its Engineering uses

26 Jather and Baiti Indian Economics 1941

27 Jain L C Indian Economy during War

28 Goulden Methods of Statistical Analysis

29 Holmes Statistics for Professionals

30 Harper Elements of Practical Statistics

31 Jones A first Course in Statistics (1921)

32 Journal of the Royal Statistical Society London.

33 Kenny Mathematics of Statistics I and II

34 King Elements of Statistical Methods

35 Kelley Statistical Methods

36 Levy and Roth Elements of Probability

37 Love Application of Statistical Methods to Agricultural Research

38 Mills C Statistical Methods as applied to Economics

39 Maclean Graphs and Statistics

40 Newsholme Vital Statistics

41 Pearl Raymond Medical Biometry and Statistics

42 Pearson Tables for Biometricians and Statisticians

43 Publications of the Board of Economic Inquiry Punjab

44 Rider Statistical Methods

45 Rhodes Elementary Statistical Methods

46 Rietz Hand-Book of Mathematical Statistics

47 Secrist, An introduction to Statistical Methods

48 Sankhya Statistical Institute, Calcutta.

49 Snedecar Statistical Methods

50 Shewart, Economic Control of Quality of Manufactured Products

51 Sorenson, Statistics for students of Psychology and Education

52 Tibbett and Crum Economics Statistics

53 Tippett The Methods of Statistics

54 Royal Statistical Society Journal

55 Statistical Publications in India

56 Waugh, Elements of Statistical Methods

57 Whittaker and Robinson Calculus of Observation

58 Yule and Kendell An introduction to the theory of Statistics

59 Zia-ud-Din, Tables of Symmetric functions for statistical purposes Proc. National Academy of Sciences 1940

60 Zia-ud-Din Research papers published in British, American and Indian Journals